

Reflections on the death of a young scientist

The tragic suicide of an outstanding graduate student at Harvard University poses troubling questions about academic priorities. But it also reinforces deeper concerns about the contemporary culture of research.

There is always something slightly humbling about human tragedy. Such is the case with the suicide two months ago of a fifth-year graduate student, Jason Altom, at Harvard University. It is impossible to speculate on the full range of factors that led an apparently outstanding and respected student, outwardly stable and well-liked by his fellow students, to decide to take his own life. What does prompt comment, however, is the fact that Altom left a detailed note describing the pressures to which he felt subject, and suggesting how some of these might have been avoided in different circumstances (see page 826).

The contents of the note, extracts of which have been published in the *Harvard Crimson*, allude to a situation with which all graduate students will be familiar. One is the constant pressure to succeed, with eyes fixed on a sometimes distant, often daunting and always challenging goal. A second is the intense relationship, which can be either supportive or destructive, with a single supervisor — a relationship that some Harvard students joke tends to last longer than most marriages.

Both pressures can be exacerbated by a lack of the financial means and social networks that might otherwise allow their more extreme impacts to be softened. Further problems are created by the system of 'indentured servitude' at some institutions, under which graduates are used to meet teaching and other commitments, and end up feeling that they are being treated as a source of cheap labour.

There is no reason to believe that the situation at Harvard, despite a hot-house culture in which many ambitious graduate students will-

ingly participate, is significantly different from that at other leading research universities. And the chemistry department, which had already been engaged in debates about mitigating such pressures, has been prompted by Altom's death to take immediate action, such as requiring every second-year student to set up a three-member pre-thesis advisory panel, and making psychological counselling services readily available.

Such moves can only be welcomed. But they inevitably raise the question, prompted by genuine concern rather than reflex recrimination, of why it took the death of an outstanding student to prompt the department into action. According to one recent PhD student, proposals for improved student oversight had been submitted by a graduate student committee three years ago, but stalled when faculty members were unable to agree on its implementation. Yet, she points out, this happened at a time when the faculty was able to conceive and start construction of a new building, renovate existing laboratories and hire new faculty.

It is impossible to pass judgement without knowing the full circumstances. But such situations raise a key issue that lies behind a broad swathe of current concerns, from scientific misconduct to the plight of contract research staff: is a culture of achievement, fanned by an increasingly competitive job market and tight competition for research grants, now in danger of driving out the culture of mutual support from which both science and its protagonists have gained so much in the past? □

Crop research meets the public

The UK government has some way to go in building trust in its handling of genetic modification in agriculture.

Lessons learnt in Britain (and no doubt elsewhere) from the BSE crisis were evident in the UK government's announcements last week of plans to change the way it regulates genetically modified crops. Out goes any lingering assumption that the technology is inherently safe, and in comes a new requirement for industry to demonstrate practically that its products will not have adverse ecological effects.

In a welcome move, a new steering group of scientists will be able to commission research it considers necessary on the ecological impacts of genetically modified crops (see page 830). The government, in turn, promises not to allow the commercialization of any crop until the scientists are reasonably satisfied that it is safe to proceed.

The government's attempts to build public trust in its scientific advice, and to include public views in its policy decisions, however, leave some questions open. It has chosen to set up a new forum of 'environmental stakeholders' whose views would contribute to decisions about genetic modification in agriculture. This forum, spanning the spectrum of interests and opinion, will need to make constructive suggestions, and avoid well-trodden, predictable and ultimately time-

wasting disputes between industry and environmentalism.

The government is already poised to embark on a survey of the public — as opposed to environmentalist — perception of the biosciences. Ministers would be wise to wait for its outcome before putting more flesh on the stakeholders' forum. But the latter should eventually be encouraged to make constructive contributions to the research agenda. In doing so, it will not only address public concerns about the risks of genetically modified crops, but should also help rebuild public trust in the application of science to foods. Here again, however, the government will need to ensure that the chance to influence research is not used as an unyielding instrument of obstruction by fundamentalist opponents of genetic modification.

Scientists on the whole are supportive of the changes, particularly the decision to authorize research trials on a commercial scale. They rightly seek better security arrangements following recent incidents of crop destruction. They should cautiously welcome, rather than oppose, increased public awareness and scrutiny of their activities while being themselves watchful (and, if necessary, vociferous) over the details of the processes the government is putting in place. □



US plans a multi-million dollar boost to computer simulation

[WASHINGTON] Universities and government laboratories in the United States are vying for leadership of a major research initiative in the scientific simulation of complex problems and other areas of computer technology.

The initiative, intended to secure US leadership in this area, will involve new expenditure of around \$200 million in its first year, rising to as much as \$1 billion a year thereafter. It is likely to be announced in President Bill Clinton's State of the Union address in January, and will form the main science and technology component of his budget proposal for the year 2000, to be presented to Congress in February.

The Department of Energy is preparing one aspect of the initiative, which will provide its non-weapons laboratories with supercomputer hardware and expertise for scientific simulation. Ernest Moniz, the under-secretary of energy, wants to give civilian scientists access to computers as powerful as those being installed at the US nuclear weapons laboratories, and has been vigorously pursuing the project since his appointment last December (see *Nature* 390, 651; 1997).

However, Clinton was told this summer that the National Science Foundation (NSF), which has no laboratories of its own and supports research through grants to universities, is best equipped to lead a major initiative in information technology — including scientific simulation. The industry-dominated President's Information Technology Advisory

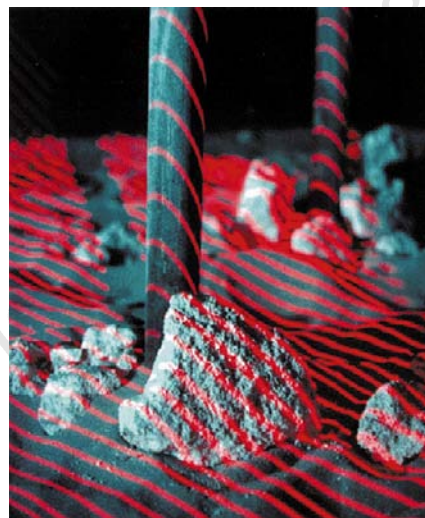
Committee (PITAC) advised Clinton that the NSF should coordinate federal computing research, and receive half the money for the new initiative.

Researchers could use the high-speed computer capability developed under the initiative to study problems as diverse as global climate change and the transport of radionuclides and other pollutants through soil and rock — an important priority at the Department of Energy, which is responsible for the clean-up of the US nuclear weapons production complex.

Structural biologists, combustion engineers and a host of other scientists and engineers interested in complex, three-dimensional problems could also use the capability.

In the next three weeks Neal Lane, the director of the White House Office of Science and Technology Policy (OSTP), will tour the facilities likely to be involved, and draw up a proposal for inclusion in the year 2000 budget; it must be submitted to the White House's Office of Management and Budget for approval by the end of November. But discussions are still taking place over how much money should be spent on supercomputers and other new facilities, and how much distributed to researchers for computer research or for developing scientific applications.

Although the proposal "is still half-baked", it will take shape over the next two or three weeks, says one White House official. In addition to the Department of Energy and the NSF, the initiative will involve the National Institutes of Health, the space



Down to detail: mapping the contents of nuclear waste tanks — as here at Hanford — is one task that is made easier by computer modelling.

agency NASA, the National Oceanic and Atmospheric Administration and the Defense Advanced Research Projects Agency.

"We're in the middle of an inter-agency discussion trying to move this forward," says Moniz. He says that, although the NSF will lead the overall initiative, the Department of Energy and NSF will have a "joint and equal" role in its simulation component.

Moniz points out that PITAC recommended extra funding of \$200 million annually over five years until the overall initiative was worth \$1 billion a year. It said this would double investment in basic research in information technology. "I'm not going to get into the numbers, but that is the scale that is being talked about," says Moniz.

Bill Madia, director of the Pacific Northwest National Laboratory in Washington State, one of the energy department laboratories set to be involved, says the simulation initiative itself will be worth \$150–\$200 million.

Government officials say the PITAC study appears to have settled any argument about whether the US computer industry needs government help with research and development. "PITAC was quite clear that fundamental innovations in computing come out of federally supported research in the universities," says one OSTP official.

However, there is some scepticism in Congress over the need for the additional spending. Past Clinton initiatives in this sphere — including the Next Generation Internet project — have been attacked on Capitol Hill as ill defined.

Colin Macilwain

Biomedicine wins British budget top-up

[LONDON] The biomedical and environmental sciences emerged as clear winners in the British government's long-awaited share-out of an extra £313 million (US\$531 million) in research funding over the next three years, announced this week.

The areas on which most of the extra funds will be spent include genome research, ageing, climate change, information technology and communications, and the social sciences.

The Medical Research Council will receive an additional £90 million between 1999 and 2002, which will mean a 6.8 per cent increase in real terms over its £290 million budget for this year.

The Biotechnology and Biological Sciences Research Council is to receive an extra £52 million, a 4 per cent increase over the same period. And the Natural

Environment Research Council gets an extra £40 million, meaning a 3 per cent increase between 1999 and 2002.

The Engineering and Physical Sciences Research Council will receive an additional £86 million over the three years, amounting to a 3.5 per cent increase. But £60 million of this must be used to support the work of the biomedical and environment research councils in areas such as bioinformatics and combinatorial chemistry.

The Particle Physics and Astronomy Research Council gets an extra £20 million, which will mean only a 0.55 per cent increase over three years. The government will reserve £30 million as protection against currency fluctuations for international subscriptions, such as that to the European Laboratory for Particle Physics (CERN) in Geneva.

Ehsan Masood

Suicide highlights graduate student woes

[BOSTON] Harvard University is reviewing its policies on graduate students following the death of Jason Altom, a fifth-year PhD candidate in the chemistry department who killed himself last August by taking cyanide.

Altom, the third Harvard graduate student to commit suicide since 1997, was working on the synthesis of a complex molecule under the supervision of organic chemist Elias J. Corey, winner of the 1990 Nobel prize.

"This event could have been avoided," Altom wrote in a suicide note, part of which was reprinted by the *Harvard Crimson*, a student newspaper. "Professors have too much power over the lives of their grad students."

Altom suggested in the note that a thesis committee, made up of three professors, should be formed earlier in the research process to help students assess their work and to protect them from what he described as "abusive research advisers". He added: "If I had such a committee now, I know things would be different."

Poor faculty advising had been highlighted last March as a serious problem in a letter to the university administration from the Graduate Student Council. Changes in the advising structure, as advocated by Altom and the student council, were incorporated in a nine-point plan adopted by the chemistry department in mid-September.

"Jason's death prompted an examination

of the role the department should play in graduate students' lives," says atmospheric chemist James Anderson, who became department chairman in July.

Under the new guidelines, each second-year graduate student will establish a pre-thesis committee composed of the adviser and two other faculty members. Students will also have "confidential and seamless access" to psychological counselling services paid for by the department, says Anderson.

Other changes include the introduction of "frequent" buffet dinners for graduate students, postdocs and faculty members, as well as regular meetings between the chairman and graduate school classes.

A survey will assess the usefulness of the new policies. "We plan to make more changes, but want to evaluate these measures first," says Anderson. The heads of Harvard's other science departments are "paying close attention to what we're doing," he adds. "They realize that the isolation and depression experienced by some graduate students is not unique to chemistry."

All 3,400 of Harvard's graduate students were questioned during the registration period in September about the effectiveness of the advisory system. "Our efforts have been galvanized by this event [Altom's suicide]," says Margot Gill, administrative dean of the graduate school. "It has forced us to ask if there is more we can do to improve

the graduate student experience."

Anne Pruitt-Logan, scholar in residence at the Council of Graduate Schools in Washington DC, endorses the idea of creating advisory committees for students. "But most of these committees focus too narrowly on the research, when students need broader mentoring," she says.

Although the advisers would not be counsellors *per se*, says Pruitt-Logan, "they would hopefully be observant enough to see that something is amiss. Having a group of advisers increases the chances that someone will catch signs of distress."

Stephen Senturia, an electrical engineering professor at the Massachusetts Institute of Technology (MIT), agrees that the adviser-student relationship is a critical issue. "It's easy to fall into the notion that the adviser owns your life," says Senturia, who runs workshops that expose students to ethical dilemmas in research and train them to handle problems with their supervisors.

Thesis committees at the university should form earlier than they do at present, he suggests, as a way of "broadening the support base for students."

But Senturia points out that MIT's Graduate Student Council is not pushing for changes in the advisory programme. That could mean that students are satisfied with the present system, he says. "Or maybe they're just too busy."

Steve Nadis

Damaged satellite adds to delays facing X-ray astronomers

[MUNICH] The German X-ray satellite ROSAT, which has been operating for more than four times longer than planned, may have to be abandoned after it was accidentally pointed towards the Sun.

The accident resulted in damage to one of its three main instruments, the high-resolution imager (HRI), developed at the Smithsonian Astronomical Observatory in Cambridge, Massachusetts. As a result, the latest call for research proposals, whose deadline was next month, has been put on ice as scientists attempt to determine the extent of the damage and assess whether a rescue is possible.

If not, the satellite's operation will not overlap with two new major X-ray astronomy missions, NASA's Advanced X-ray Astrophysics Facility (AXAF) and the European Space Agency's XMM. An overlap would have kept new data flowing to the X-ray astronomy community without a break.

ROSAT was launched in summer 1990, with an intended lifespan of 20 months. One of its main goals was to carry out an all-sky survey of X-ray sources, and a catalogue of

such sources was published by the Max Planck Institute for Extraterrestrial Physics in Garching, Germany, in 1996.

As well as the HRI, ROSAT's other main instruments are a Position Sensitive Proportional Counter and a Wide Field Camera. After the satellite's positioning system failed last April, attempts were made to adapt the Wide Field Camera, developed

at the University of Leicester, to take over the imager's position-sensing function. But a chain of events led to a total loss of positional control at the end of September.

The scientists say there is little chance of reviving the HRI.

Funding for the satellite's operation has been guaranteed by the German ministry of research until the end of 1999, but none of ROSAT's instruments is likely to be usable for that long.

One consequence of the technical problems is that the life of the third main instrument, the Positional Sensitive Proportional Counter, has been reduced to a few days. This instrument measures both

the position and energy of X-ray sources, and was developed at the Max Planck Institute for Extraterrestrial Physics.

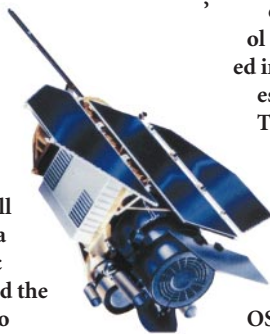
The Wide Field Camera has already completed most of the viewing within its technical capability, says Martin Ward, director of the University of Leicester's X-Ray Astronomy Group.

But X-ray astronomers are facing additional disappointments. The launch of AXAF, originally planned for last August, was delayed until next January following technical problems, and last week a further delay of several months was announced. XMM has been delayed for six months, until February 2000 (see *Nature* 395, 732; 1998).

Ward says that, although the likely loss of ROSAT is unfortunate, there are still opportunities for X-ray astronomers to do new work, for example with the Italian mission SAX, launched in 1996.

Joachim Trümper, the director of the ROSAT project, is disappointed but pragmatic. "We expected two years," he says, "so with a happy and fruitful life of eight years we can't complain. We knew it would have to end sometime."

Alison Abbott





Zecchino: under pressure to revise reform decree.

Italian scientists fear impact of cabinet reshuffle on reforms

[MUNICH] Italian scientists are reacting nervously to the appointment of Ortensio Zecchino, a lawyer with no experience of scientific issues and a member of the right-wing PPI (Partito popolare italiano), as minister for research and universities in the new coalition government.

Last week's ministerial reshuffle followed the collapse of the administration of former prime minister Romano Prodi after losing a no-confidence vote. Zecchino's predecessor was Luigi Berlinguer, a member of the left-wing PDS (Partito democratico della sinistra), who had had broad responsibilities for education. He remains in the cabinet, but his remit has been limited to schools.

The main concern of scientists is the effect the government's fall will have on the major restructuring that Italian research has been facing to increase efficiency (see *Nature* 386, 208; 1997).

The broad outline of the reforms was approved earlier this year. But pressures of other government business meant that decrees outlining rules for individual organizations — in particular the Italian space agency, the energy and environment agency, and the national research council (CNR) — had not been formally approved when the government collapsed.

One of Zecchino's first tasks will be to ensure that the decrees are approved and implemented. But he has the power to make changes to them before this happens, and many researchers are keen that he should introduce revisions to the decree covering the CNR.

A draft of the CNR decree had angered university professors as it eliminated their power over the research council's policy-making and grant distribution. Researchers were also worried that it concentrated that power in the hands of a small group of political appointees (see *Nature* 394, 712; 1998).

Lucio Bianco, president of the CNR and a member of the same party as Zecchino, is keen that the council should not lose its previous autonomy.

Alison Abbott

Alarm raised over drop in basic research at NASA

[WASHINGTON] Support provided by the US space agency NASA for basic space science research has declined dramatically in the 1990s, according to a report published last week by the National Academy of Sciences' Space Study Board.

The decrease is due partly to a shift towards smaller, cheaper missions that no longer include the cost of data analysis in their budgets. "As the faster-paced style of the agency has begun to take hold," the report says, commitment to research and data analysis (R&DA) may be getting "blurred or even lost".

The report also finds little evidence of space research moving from NASA laboratories to universities, as was recommended in a report published three years ago.

A panel chaired by Anthony England, an Earth and space scientist at the University of Michigan, spent two years looking at trends in NASA's \$1.5 billion R&DA budget, and found some "alarming results".

Total NASA research spending grew by 44 per cent from 1991 to 1998, and remained at 3 per cent of the agency's budget. But the numbers are deceptive, as much of the increase was spent on the Earth Observing System satellite data system and on technology development in the space science office.

Meanwhile, 'traditional' research and analysis — grants for university and NASA scientists to do basic research, interpret spacecraft data, and build new instruments — decreased by 22 per cent over the period.

Although funding for NASA grants and the number of grants both went up, the size of a typical award decreased by 25 per cent from 1986 to 1995. Hardest hit were disciplines such as astronomy, space physics and planetary science, for which the median grant dropped from \$64,000 to \$59,000.

Wesley Huntress, director of the Geophysical Laboratory at the Carnegie Institution of Washington, and until recently the head of NASA's space science office, admits that grant awards have become "too damn low".

The median size of Earth science grants remained level from 1986 to 1995. But one group — the life science and microgravity researchers who design experiments for the space shuttle and space station — saw their median grant rise from \$69,000 to \$100,000.

One reason for this rise is that shuttle experimenters often use funding to build expensive flight hardware as well as to conduct research. But life science and microgravity researchers may also have fared better than their colleagues, says England, because "most of NASA's future is tied up in the space station" and the agency has made a point of nurturing that research community.

Other agency-sponsored research does not seem to have the same support. The long-standing criticism that NASA is more interested in flying spacecraft than in doing science remains even more of a concern in the age of cheaper, faster missions.

In the past, large projects, such as Voyager or the Galileo Jupiter mission, dedicated a portion of their budgets for scientists to analyse the data they produced. With less money and shorter lead times, flight projects now barely have the resources to produce raw data, and generally leave the interpretation to NASA's R&DA programme.

Unlike the more visible and politically saleable flight projects, R&DA's 'softness' makes it vulnerable to budget cuts, says the academy panel.

The report also addresses the balance between in-house NASA research and extramural grants to universities, which England calls a "very sensitive issue". Even though a 1995 academy report recommended that most scientific research should be conducted outside the agency, researchers at NASA field centres still get most grant money.

NASA scientists receive 40 per cent of the R&DA budget, and another 30 per cent goes to industry scientists, many of whom are on-site NASA contractors. Only 26 per cent of the money goes to universities.

Reduced NASA funding has already affected the thousands of researchers who depend on agency grants. England says space science is "certainly not a growth field", and may not even be a stable one. The report recommends that NASA should make greater use of training grants for graduate students and "rotate" more outside scientists to serve temporary duty at the agency.

But the panel took pains to emphasize that their report was "not an appeal for an increase in NASA funding". Rather, they say that more attention — including regular outside peer review — should be paid to the balance between flight projects and the science that supports them.

England says the committee had a difficult time finding budget figures that clearly showed NASA's research priorities. The panel had to hire a former NASA budget analyst to help make sense of the numbers. This led to the panel saying that, if NASA cares about its R&DA programme, it needs to do a better job of tracking how money is spent.

"The fragmented budget structure for R&DA makes it difficult for the scientific community to understand the content of the programme and for NASA to explain the content to federal budget decision-makers," says the panel.

Tony Reichhardt

US panel split on endocrine disruptors...

[WASHINGTON] The US Environmental Protection Agency (EPA) is pressing ahead with a plan to screen 15,000 chemicals for their possible effects as endocrine disruptors in animals and humans — without recourse to what some claim to be adequate independent scientific advice.

The situation has arisen because of delays, reportedly due to disagreements between experts with conflicting views, in the publication of a long-awaited report from the National Research Council (NRC). According to NRC officials, the report, which is intended to provide a definitive assessment of endocrine disruptors for the US government, is not now likely to appear until early next year.

An EPA spokeswoman says the environmental agency used its own scientific assessment as the basis for its actions, which could cost hundreds of millions of dollars.

But scientists and officials on both sides of the issue view the non-appearance of the NRC study with alarm. "My feeling is you need to assess the science before you proceed with the screening," says one official.

The NRC, the operating arm of the National Academies of Sciences and Engineering, was set up to provide definitive advice to the government on difficult technical questions. Few problems are tougher or more contentious than the proposed regulation of endocrine disruptors. These are chemicals that, according to the endocrine-disruptor hypothesis favoured by some scientists, can induce a variety of health effects in humans and wildlife by mimicking or interfering with the actions of hormones.

Chemicals identified as endocrine disruptors include some pesticides, industrial chemicals, drugs and contaminants. But industry groups and other scientists say there is no hard evidence linking any of

these to human health problems.

The EPA and other government agencies are paying \$860,000 for the NRC study, which was commissioned in March 1995. The panel first met in October 1995, and was expected to complete its work by the spring of 1997. But that deadline soon slipped to spring of this year, and then to this autumn.

Most of those following the issue now expect publication in December. But Jim Reisa, director of the board on environmental studies and toxicology at the NRC, which is overseeing the study, says he is "not wildly optimistic" that it will appear before 1999.

The NRC panel has been dogged from the beginning by bitter disagreements about the endocrine-disruptor hypothesis. Members of the panel include noted opponents of the hypothesis, such as Stephen Safe of Texas A&M University and James Lamb of Jellinek, Schwartz & Connelly, an industry consultancy based in Washington, as well as researchers who believe their work has verified it, such as Frederick vom Saal of the University of Missouri at Columbia.

The delay has been "mostly due to the difficulty in reaching a consensus among the panel members", says Reisa. "We didn't go into this blindly. We appointed members with strong views on both sides, in the hope that, if we reached a consensus, it would carry great weight. I still hope that we will reach consensus."

According to Reisa, the current version of the report has just come back from 17 external reviewers, who have provided 120 pages of comments for the panel to argue about.

Panel members will not comment on the substance of their disputes. But sources close

to some of them say the main draft is cautious in its assessment of the extent to which endocrine disruptors pose a risk to human health.

Four environmentally inclined members of the panel, including vom Saal and Ana Soto of Tufts University, are said to be holding out against the draft. They have submitted their own alternative executive summary of the report, which may end up as a minority appendix attached to the main study. The NRC tries to avoid issuing minority reports, but occasionally does so.

The conflict has left supporters of the endocrine-disruptor hypothesis playing down expectations of the study, which was supposed to provide the scientific basis for US government policy. Pete Myers, for example, director of the W. Alton Jones Foundation in Charlottesville, Virginia, suggests that the impact of the report "will depend on whether it is able to rise above the politics" of the issue.

"The report has suffered because of the willingness of some people on the panel to engage in political debate" during the course of the study, argues Myers. He accuses Safe and Lamb, in particular, of such activity.

Two weeks ago Lamb took part in a press conference in Washington at the Society of the Plastics Industry to release an unpublished study, sponsored by the plastics industry. The study refutes a recent finding by vom Saal that bisphenol A, a chemical used in plastics production, caused prostrate growth and low sperm counts in mice (*Toxicology and Industrial Health* **14**, 239; 1998).

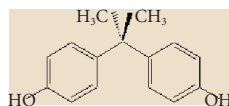
Lamb said he had reviewed the unpublished study, and found its methodology superior to vom Saal's.

Perhaps the only aspect on which the two sides of the argument agree is that the delays in the NRC study will diminish its impact. The EPA, responding to laws passed in 1996, established its own Endocrine Disruptor Screening and Testing Advisory Committee, which last month recommended that 15,000 chemicals should be screened and, if necessary, tested on animals for health effects.

Earlier this month, the EPA accepted its recommendations and said it would prepare a screening process by March of next year. According to scientists on both sides of the argument, the NRC report was originally supposed to provide the scientific basis for this process.

But the train of endocrine-disruptor regulation is now leaving the station, raising doubts about the usefulness of the NRC's meticulous study process in the context of fast-moving policy debates. Reisa can suggest only one way in which it could have been speeded up: "We could have picked an easier subject!"

Colin Macilwain



Bisphenol A: one of the suspect chemicals.

hypothesis, such as Stephen Safe of Texas A&M University and James Lamb of Jellinek, Schwartz & Connelly, an industry

consultancy based in Washington, as well as researchers who believe their work has verified it, such as Frederick vom Saal of the University of Missouri at Columbia.

The delay has been "mostly due to the difficulty in reaching a consensus among the panel members", says Reisa. "We didn't go into this blindly. We appointed members with strong views on both sides, in the hope that, if we reached a consensus, it would carry great weight. I still hope that we will reach consensus."

According to Reisa, the current version of the report has just come back from 17 external reviewers, who have provided 120 pages of comments for the panel to argue about.

Panel members will not comment on the substance of their disputes. But sources close

... while Japan studies drop in sperm counts

[TOKYO] Japan's Ministry of Health and Welfare announced last week that it is to sponsor a nationwide study of sperm counts. The move follows public concern over the possible effects of endocrine disruptors — man-made chemicals suspected of disrupting human reproductive functions — on male fertility (see above).

Researchers from St Marianna University School of Medicine, Sapporo Medical College, Kanazawa University and Osaka

University plan to begin the project later this year.

According to the health ministry, semen samples will be collected from 1,500 men from all over Japan. The samples will be tested to determine the number of sperm per millilitre and the percentage of sperm with normal motility. The researchers also hope to determine whether certain types of food have any effect on sperm levels.

Widespread concern about the effects of

endocrine disruptors was triggered last year by the release of a report by the Environment Agency, which listed 67 chemical compounds, such as dioxins and polychlorinated biphenyl compounds (PCBs), that are suspected of mimicking natural sex hormones.

The government has since allocated ¥18 billion (US\$152 million) for research into the effects of such compounds on human health (see *Nature* **392**, 748; 1998).

A.S.

French inquiry into misconduct is shelved

[PARIS] The French government has quietly dropped plans to commission an international inquiry into allegations of misconduct at a laboratory of INSERM, the national biomedical research agency (see *Nature* 393, 203; 1998).

In May, Daniel Nahon, the director-general of the Ministry of National Education, Research and Technology announced that four international experts would be asked to carry out an inquiry into the activities of the INSERM Laboratory of Nutrition, Lipoprotein Metabolism and Atherosclerosis at the University of Rennes 1, and its director Bernard Bihain (see *Nature* 391, 519–520 & 825; 1998). He promised that the inquiry would be completed within three months and its conclusions made public.

The laboratory itself was closed by INSERM in August on the grounds that Bihain had asked to be discharged from his duties on “personal grounds”. At the time, many scientists expressed concern that the closure might be used to justify a reduction in efforts to get to the bottom of the affair.

Their fears appear to have been realized.

Nature has learnt that the ministry has now abandoned the inquiry. “We stopped everything,” admits Vincent Courtillot, principal adviser to Claude Allègre, the science minister. “The affair is over, the guy is in America, the lab does not exist, so we are not going to make an inquiry committee work on something where there are no consequences.”

Jacques Lenfant, vice-chancellor of the University of Rennes 1, says he is deeply disappointed by the government’s action, and questions why it began an inquiry in the first place. The scientific board of INSERM has also repeated its demand that a scientific investigation of the laboratory be carried out and made public “in the interests of the agency and its staff”.

For its part, the SNTRS-CGT, a trade union representing INSERM scientists, has called for “full light to be thrown on the presumptions of scientific fraud”, arguing that “the honour of the scientific community is at stake”.

In a thinly veiled allegation that the government has tried to hush-up the affair, it asserts that “only the decision-makers had

to be worried about the outcome of a search for the truth”.

Meanwhile, a separate inquiry into the activities of the laboratory by the police fraud squad in Rennes (see *Nature* 394, 308; 1998) has moved a step closer to becoming a formal judicial inquiry. The public prosecutor of the region has opened a preliminary investigation into the affair, following the submission of an ‘information file’ (*dossier de renseignement*) by the fraud squad making the case for a judicial inquiry.

INSERM, which has been widely criticized over its handling of the Bihain affair, is to create an office of scientific integrity. Claude Griscelli, the agency’s director-general, says the unit’s remit will be to investigate “promptly and effectively” all allegations of misconduct, and to settle disputes, for example over the order of authorship on papers.

In a bid to prevent future misconduct, the unit will also make recommendations for good laboratory practice. These would include guidelines for clinical trials and standards for keeping laboratory note-books and other records.

Declan Butler

Controversial Swedish science minister loses seat in reshuffle

[STOCKHOLM] A cabinet reshuffle in Sweden has led to the demotion of its controversial science minister, Carl Tham, who had angered many influential members of the scientific community with sweeping changes to higher education and research.

In his four years as science minister, Tham redirected funding from basic to applied research, with increased emphasis on the interaction between science and society, including closer cooperation between science and industry.

Tham had also expressed a desire to reform power structures within the academic research community, breaking down interdisciplinary barriers and giving minority groups greater access to both higher education and research.

Tham lost his post in a reshuffle earlier this month following a general election in which support for the ruling Social Democrats fell by 11 per cent, although they continue to govern with support from the former communist party and the Greens. Tham has been replaced by Thomas Östros, the former tax minister.

No official reason has been given for the change. But Tham had faced a growing band of critics as he moved to increase political influence over the allocation of resources and decisions about research priorities, challenging the traditional autonomy of the universities.

The traditional academic hierarchy also

opposed Tham’s reforms to substantially increase the numbers of professors at universities by giving this title to former senior lecturers.

Tham also established several new, regional universities, reallocating money to them from the established universities, and gave a number of smaller and regional universities the status of ‘complete’ universities.

Furthermore, 30 new professorships in various subjects have been created, earmarked for specified groups, particularly women. A number of new posts in ‘gender research’ were also established, and reforms introduced to speed up the passage of students through higher education.

Tham’s reforms have met with considerable support from those who have benefited from them, such as the feminist movement and representatives of the regional universities. But they have also generated repeated protests from large sectors of the scientific community.

Shortly before last month’s general election, for example, more than 300 professors of medicine, natural sciences and technology signed a public protest against the reallocation of money from basic to applied research.

“What has taken decades to build up is about to be lost,” says Dan E. Nilsson, professor of zoology at the University of Lund, who initiated the protest. “We need



Tham (centre), seen here on an official visit, has faced growing criticism of his radical policies.

quick and forceful reactions to keep Sweden at the forefront of science.”

Dan Brändström, head of The Bank of Sweden Tercentenary Foundation, one of the country’s largest independent sources of science funding, says it is important to rebuild confidence between science and politics. “Diplomacy, dialogue and respect for the problems that science is facing during a time of restructuring is what we hope for from the new minister,” he says.

But whether the appointment of the new minister will lead to any real change in policy remains to be seen. Sweden’s prime minister, Göran Persson, has declined to make any public criticism of his former science minister. This suggests that, even though Tham has left, the policies he introduced may remain in force.

Peter Sylwan

UK holds up applications of genetically modified crops

[LONDON] The British government and life-science companies have agreed a temporary moratorium on the commercialization of genetically modified crops, pending research into their ecological effects.

The moratorium is part of a comprehensive package of measures announced last week that are designed to allay public fears about genetic modification in agriculture.

Under the measures, herbicide-tolerant crops will be withheld from the market for at least a year, and insect-resistant crops for three years, while research is carried out into their impact on biodiversity, for example.

A steering group of scientists will be asked to identify relevant research questions and to plan experiments to answer them. Industry is expected to contribute to the costs of the experiments, which will be monitored by independent experts. Many of the experimental trials will be on a commercial scale.

The government's conservation advisory body, English Nature, will help to set the agenda, and non-governmental organizations will be encouraged to play a "constructive role", says a government spokesman.

Michael Meacher, the environment minister, said last week that the government is "right to be cautious" and will "make sure that for every product we have practical evidence on safety before we can take a decision to move to commercialization".

The measures include a review of the use and ecological effects of herbicides and pesticides on genetically modified crops. Companies wanting to grow commercial-scale genetically modified crops must show that there are no adverse ecological effects.

The government wants this requirement to be included in the directive governing the commercial release of genetically modified crops in all other European Union states.

It also plans to set up an 'environmental stakeholders forum', which will allow those

with an interest in genetically modified products to have their views taken into account when decisions are made.

The forum is a response to calls for greater public representation on the government's scientific Advisory Committee on Releases to the Environment, which advises the government on whether or not a genetically modified crop is safe to grow. While this committee will remain science-based, it will consider the forum's views when making recommendations to government.

The government is also to set up a ministerial committee on biotechnology, which will comprise junior ministers from ten government departments.

A government spokesman says that the committee's first tasks will be to set up the forum, and to review the complicated regulatory structure governing the transfer of genetically modified organisms from the laboratory to the supermarket. One option is a proposal from the Royal Society to set up a single body to oversee the work of the different regulatory advisory committees (see *Nature* 395, 5; 1998).

The measures have had a mixed response from critics of the government's previous approach to genetically modified crops. These critics, which include environmentalist groups, organic farmers and even English Nature, had opposed the commercial-scale planting of such crops while questions about their ecological impact remain unanswered.

English Nature considers the government's decision to be "wise". Brian Johnson, the body's adviser on genetically modified organisms, says it is good news that industry must demonstrate that genetically modified crops have no adverse ecological impact.

Conservation agencies, he says, have been concerned for a while that the use of genetically modified herbicide-tolerant and insect-resistant crops could greatly reduce weeds and insects on farmland, threatening the survival of several species of farmland birds.

"At the moment, a company is asked simply to state, 'yes' or 'no', whether its genetically modified crop will affect other organisms," says Johnson. He says companies invariably answer 'no' even though there is no proof.

Most environmentalist groups remain critical of the measures, which are unlikely to stop protesters digging up field trials of genetically modified crops (see *Nature* 394, 608; 1998). Far from limiting genetically modified crops, they say the government's decision to insist on commercial-scale experimental trials amounts to even greater release of a product that most consider to be an environmental pollutant. **Ehsan Masood**

Canada announces second round of infrastructure awards

[MONTREAL] The Canada Foundation for Innovation last week announced research infrastructure awards worth Can\$21.6 million (US\$14 million) to help strengthen the country's capability for research and technology development.

More than 550 researchers at 35 centres will benefit from the awards, which are being made under the Institutional Innovation Funds and Regional/National Facilities, and the Research Development Fund.

This is the second such announcement this year. The first awards, in the foundation's New Opportunities programme, totalled Can\$36 million. They were designed to help young researchers in their first academic appointments to obtain facilities (see *Nature* 394, 712; 1998).

There has been an enormous response to competitions for the awards. In the current series, more than 450 applications were submitted by Canadian universities, hospitals and non-profit institutions. The awards targeted key needs in health, science, engineering and the environment.

A striking aspect of the competition was that many institutions submitted joint proposals. "What has surprised us is how people from different disciplines have been eager to work together on projects," says Susanne Fortier, vice-president of research at Queen's University at Kingston, Ontario.

But, for the 159 projects approved after an initial review, even more cooperation is going to be necessary. Because of their complexity, their proposers will be invited to submit revised projects for further review before funding decisions are made in 1999.

David Strangway, president of the Canada Foundation for Innovation, cites as an example proposals from institutions and groups of institutions for what he calls a digital library.

"You have to work out the site licensing with the publishers... to buy a single site licence which can serve all the university libraries. So we're going to have those groups come back to us with a national proposal."

Another complex area was high-performance computing, Strangway said. "We have a lot of proposals from institutions. They will have to show how they are going to provide access to all institutions across the country. There were also outstanding proposals in genome studies."

The detailed strategy for funding these more complex projects will be announced shortly, he said. Phase two will be extremely competitive because projects worth a total of Can\$735 million are competing for funding of only Can\$370 million. **David Spurgeon**



Brazil forced to cut back science funding

[LONDON] The Brazilian government is expected to announce major cuts to the country's science budget this week — including a block on all new fellowships and research grants — as part of its response to the international financial crisis.

A broad package of fiscal measures has been drawn up following talks between the government and the International Monetary Fund, which are due to culminate with a set of economic measures designed to help Brazil weather its approaching economic recession.

Science and technology have not been spared from a wide swath of cuts to public spending. The Ministry of Science and Technology is being required to make cuts of at least 248 million réals (US\$208 million) in this year's budget, 20 per cent of its planned total of R\$1.2 billion.

The National Council for Scientific and Technological Development (CNPq), the main research funding agency, is expected to see its budget reduced by about R\$50 million from the current level of R\$436 million.

Oskar Klingl, chef de cabinet to Brazil's science minister, Israel Vargas, says the difficult situation facing science is now unavoidable. The budget cuts for 1998 are already known, but uncertainty still hangs over what is likely to happen next year.



Brazil's Finance Ministry was under repair last week as it moved to meet the IMF's demands.

In addition to the cuts at the CNPq, other budget reductions will have a direct impact on the activities of the 14 research institutions run by the science ministry.

The situation has spread alarm in the scientific community. Last week a public protest in Rio de Janeiro led by the Brazilian Society for the Development of Science (SBPC) brought together scientists representing more than 30 organizations.

Their main concern is for the future of the CNPq, the main source of funding for university research. Two weeks ago, the CNPq circulated internally a list of proposed economy measures, signed by its president, Jose Galizia Tundisi, that included freezing the approval

of practically all new fellowships, research grants and cooperative agreements.

A senior CNPq official says the measures have been carefully considered, and that the agency tried to ensure that the cuts did the minimum damage. All formal commitments will be met, and fellowships that have already been granted will be funded.

But this has done little to reassure scientists. Last week's protest culminated with the release of two statements, one directed to the Brazilian nation and the other to the international scientific community, attacking what the protesters described as the government's lack of interest in science and technology.

Scientists have called for Tundisi's resignation, complaining that he failed to discuss the agency's financial difficulties with the scientific community, but merely drew up the internal list of economy measures. The protests were triggered by the 'leak' of this list.

Many scientists fear that their position will become even more difficult after the release of this week's austerity package. The SBPC has created a forum on its website for discussion of the topic. Messages of support have been received from scientists in other Latin American countries, many of whom share similar concerns about their own countries' science budgets. **Andrea Kauffman-Zeh**

Clash over demand for more synchrotron sources in Europe

[PARIS] Simmering frustration among French synchrotron researchers over the government's reluctance to finance the construction of a third generation facility boiled over this week. Staff at the national synchrotron facilities accused the science minister, Claude Allègre, of making misleading statements to the National Assembly on the need for a new facility.

Plans for a FF1 billion (US\$182 million) 2.15-GeV source, Soleil, to replace the ageing LURE facilities near Paris (800 MeV and 1.85 GeV), have received backing from Catherine Bréchnignac, director general of the Centre National de la Recherche Scientifique and Yannick d'Escatha, head of the French Atomic Energy Commission (see *Nature* 390, 212; 1997). But the scheme has been frozen by Allègre as part of a review of big science spending.

The French proposal was 'strongly endorsed' in a report on European synchrotron radiation, issued last week by an independent panel of the European Science Foundation. The report warned that, unless new sources were built, Europe "risked losing its lead in certain aspects of biological and biomedical research".

Synchrotron radiation is used by scientists from many disciplines to reveal

the structure of biological molecules. The panel points out that demand will increase as the plethora of genome projects creates a need to determine more protein structures.

But in the French National Assembly last week Allègre argued that Europe had a surplus of synchrotron facilities. "Within two years there will be seven third-generation synchrotrons in Europe," he said. "There are two in the United States. Do you really think we need an eighth without a preliminary inquiry? Do you know an area where Europe needs four times as many machines as the United States?"

Allègre reiterated his view that most big science facilities should be European, and that national facilities should be the exception. He also asserted that Soleil would provide employment for only 40 people, and that the construction of synchrotrons in Germany and Italy would free user time at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France.

Allègre spoke in response to questions from Pierre Lasbordes, a member of the neo-Gaullist RPR party, whose Essonne constituency is home to LURE and a candidate to host Soleil. Lasbordes said that the Paris region agreed last year to pay more than half the costs of a new machine.

"We are only waiting for you," said Lasbordes, describing Soleil as "indispensable". The delays were "all the more regrettable" given that new facilities were being built in several European countries, including the United Kingdom.

LURE's director Robert Comès and senior staff at the facility have written to Allègre, claiming that his declaration to parliament was inaccurate, and showed his "lack of knowledge" of the project. They claim the United States has plans to increase its number of synchrotron sources to seven.

The researchers contest the need for a further inquiry into the case for Soleil, arguing that it has been approved by several assessments over the past seven years. The letter challenges Allègre's estimate of employment at the facility, pointing out that LURE itself employs 350 full time staff, and is used by almost 2,000 researchers annually.

It also challenges Allègre's claim that the refurbishment of Italian and German facilities will free user time at ESRF, adding that soft radiation produced by Soleil will find different applications to the hard X-rays from the 6-GeV ESRF machine. The letter also points out that France's share of user time already exceeds its financial participation in the ESRF. **Declan Butler**

Online chemistry journal slashes publication costs

[LONDON] Libraries in the United States and Europe have joined forces with the UK Royal Society of Chemistry to provide an electronic peer-reviewed journal at a cost more than 20 times less than some printed rivals. The journal, *PhysChemComm*, is the result of an initiative between the society and the Scholarly Publishing and Academic Resources Coalition of the United States, aimed at bringing down the cost of journals and reducing the time taken to publish articles (see *Nature* 393, 719; 1998).

Participants also include the UK Standing Conference of National and University Libraries, and university libraries in Denmark. *PhysChemComm* will be available next year at an annual cost of US\$353 to institutions. This compares to \$8,000 for a comparable existing title.

Report slams medical training in Austria

[MUNICH] Austria's medical students have little contact with patients and almost no experience of research, according to a report from an evaluation committee

commissioned by the Austrian Rectors' Conference. Research training is particularly important because many students use the basic medical programme as training for research careers in the life sciences.

The report is based on a self-assessment exercise completed by three medical faculties, and visits to some of their institutes. Austria's medical education is in dire straits. Of the 3,000 students a year who embark on medical studies, half drop out within two years.

Science advisory panel starts work in France

[PARIS] The inaugural meeting of France's National Science Council was held last week to advise the government on national research priorities and science policy. The 27-member committee is attached to the office of the prime minister, Lionel Jospin, and chaired by Claude Allègre, the research minister. It will meet every six months.

Committee members include nine scientists from other European countries, such as Fotis Kafatos, director of the European Molecular Biology Laboratory in Heidelberg, Germany, and Nicolas de Voogd, president of the Delft University of Technology in the Netherlands. Seven industrialists have also been appointed, including the genome pioneer, Daniel

Cohen, director general of Genset, the French biotechnology company. French scientists on the committee include Nobel prizewinner Claude Cohen-Tannoudji.

Cancer specialist takes top position at FDA

[WASHINGTON] Jane Henney, a 51-year-old cancer specialist who has been vice-president for health sciences at the University of New Mexico since 1994, was confirmed last week by the US Senate as the new administrator of the Food and Drug Administration.

The move came after Senator Don Nickles (Republican, Oklahoma) withdrew objections based on his suspicion that Henney would work too hard to regulate tobacco products, and might allow the 'abortion drug' RU-486 to be sold in the United States (see *Nature* 395, 632; 1998).

Russia beats silicon blockade of India

[NEW DELHI] India is to receive silicon wafers for its civil and defence semiconductor industries from its old ally Russia. The move is an attempt to circumvent the ban on their export from Western countries as part of sanctions for conducting nuclear tests. The joint venture with Russia, however, dates

from 1993, and has been revived under a science agreement between the countries.

Single crystal silicon ingots will be manufactured at one of Russia's dismantled nuclear facilities in Siberia. The ingots will be converted into polished wafers in India.

Dutch launch for giant radio telescope network

[LONDON] A network of 16 radio telescopes that will effectively create a radio telescope comparable in size to the diameter of the Earth was inaugurated in the Netherlands last week. The network, known as the European Very Long Baseline Interferometer (VLBI), links up radio telescopes from Europe, the Ukraine and China.

The VLBI will have a resolution of fractions of one-thousandth of a second of arc, enough to read this page from several thousand kilometres away. Its heart is effectively a purpose-built supercomputer capable of handling 16 billion bits per second from the combined individual telescopes.

Mir space mirror could cut Earth's power bills

[WASHINGTON] A space mirror sent to the Russian Mir space station last week will test future schemes for illuminating light-starved

northern cities from orbit, as well as space power and propulsion applications. The 2.5-metre thin-film 'Znamya' mirror will be unfurled from an automated Progress cargo ship in February, after it leaves the Mir station.

At times during the day-long experiment it should be visible from some northern cities as a moving spot with the brightness of five to ten full moons. A spokesman for RSC Energia, one of the Russian sponsoring organizations, says: "This is just one step towards space illumination. It remains to be seen whether this is a practical application."

Central American support for Moncada protests

[LONDON] The universities of El Salvador and Honduras have joined the protests against the decision not to include Salvador Moncada among the winners of this year's Nobel prize for physiology or medicine.

Moncada was widely tipped as a strong candidate for the prize, which was awarded to three US scientists for work on nitric oxide as a signalling molecule in the cardiovascular system (see *Nature* 395, 734; 1998). Both universities will officially protest to the Nobel committee this week. The Universidad De El Salvador, where Moncada studied medicine, has allocated space on a website to

the protest, and is inviting readers to "denounce this act of injustice" (<http://www.ues.edu.sv/>). Moncada is currently director of the Wolfson Institute for Biomedical Research at University College London.

How Zeiss put stars in everyone's eyes



[MUNICH] Seventy-five years ago, the Jena-based optical company Carl Zeiss introduced a technology to bring a reliably starry — and always cloudless — sky to the public. Zeiss's planetarium projector (above) was displayed at the Deutsches Museum in Munich on 21 October 1923. Two years later, the Deutsches Museum, the world's largest museum of science and technology, became home to the world's first public planetarium.

DEUTSCHES MUSEUM

Rising costs hold up drug discovery

Sir—Sally Lehrman speculates that the volatile and declining US stock markets may eventually “trigger more investor interest” in biotechnology stocks (*Nature* **395**, 104; 1998). The fundamentals of the biotech industry argue otherwise: there are too many companies pursuing too few products that are fantastically expensive to move through the regulatory pipeline.

The regulation of drugs has become so burdensome in the United States that the time required for development — from discovery to market approval — has more than doubled since 1964, from 6.5 to 14.8 years. During the 1990s the time from the start of clinical trials of a drug to its marketing approval has been lengthening. Such prolonged research demands an ever-increasing amount of capital. From 1977 to 1996, approvals of new chemical entities by the Food and Drug Administration (FDA) remained relatively flat, while research

spending by pharmaceutical companies increased from \$3 billion annually to almost \$20 billion.

Bringing a drug to market in the United States now costs more than \$500 million, by far the highest price tag in the world.

Biotech companies have enjoyed less than stunning success at jumping through the FDA's hoops. According to the FDA's website (www.fda.gov), the agency approved only two new biotech drugs in 1994, one in 1995, none in 1996, five during 1997, and none during the first quarter of this year. (These figures do not include duplicates of products already marketed by other companies.) That dubious record is reflected in the performance of biotech mutual funds, which, according to Charles Schwab & Co. (www.schwab.com), have consistently underperformed the Standard and Poor's 500 (a major stock index) during the past ten years.

These regulatory costs and delays have given rise to an inauspicious imbalance between products and companies. In April the Pharmaceutical Research and Manufacturers of America organization counted only 350 biotech-derived drugs in clinical trials from the 1,000 US biotech companies. Drug developers cannot afford to gamble on products that are not likely to be blockbusters.

Although biotechnology applied to pharmaceuticals has made signal contributions to medical therapeutics, it languishes far behind its potential. Life-saving products will continue to emerge, albeit at a trickle of what is possible, but from the vantage point of most companies and investors, there seems little reason for great optimism.

Henry I. Miller

*Hoover Institution, Stanford University,
Stanford, California 94305-6010, USA*

Gender gap in health decline in East Europe

Sir—The health decline among middle-aged men in Eastern Europe between 1950 and the late 1980s associated with cardiovascular disease¹ has continued. Between 1990 and 1994, male life expectancy in Russia, for example, fell by more than six years², yielding the widest gender gap in life expectancy (13.5 years) of all European countries. Intervention efforts aimed solely at the traditional risk factors of smoking, blood pressure, lipids and obesity are insufficient to stop this epidemic. A more productive approach to prevent premature death among Eastern European men might be to strengthen social relationships, decrease social isolation and depression, and to increase adaptive coping skills.

Eastern and Western European men differ very little with regard to standard coronary risk factors, their alcohol consumption and obesity³. However, there are striking differences between Eastern and Western European men in psychosocial risk factors: for example, Lithuanian men, who are four times more likely to die from coronary heart disease than Swedish men⁴, report more depression and exhaustion, less social support and integration, and fewer effective coping strategies than their Swedish counterparts⁵.

Women in Eastern Europe have not been affected in the same manner by the cardiovascular disease epidemic². They also smoke less, are less likely to be hypertensive,

and have higher levels of protective high-density lipoprotein cholesterol than men. Even so, only 40% of the variation in the gender ratios of mortality from coronary heart disease in 24 countries (including Russia, Lithuania, Poland and the former Czechoslovakia⁵) can be explained by differences in the known risk factors. Women may be less affected because they report more social support, are more socially integrated, and cope better with life crises and disasters than men^{6,7}. Men are more likely to use avoidant coping, such as denial, distraction and increased alcohol consumption, whereas women use vigilant strategies⁶. Although women report more depression, they are more likely to accept depression as a disorder to be treated than are men⁸.

These gender differences may contribute to the health decline among Eastern European men, who have to cope with economic uncertainty, disruption of traditional male roles and the break-up of social relations, as well as the social stigma associated with the need to ask for help or turn to one's social network for support.

Gerdi Weidner

*Department of Psychology,
State University of New York at Stony Brook,
Stony Brook, New York 11794-2500, USA*

1. Feachem, R. *Nature* **367**, 313–314 (1994).
2. Notzon, F. C. et al. *J. Am. Med. Assoc.* **279**, 793–800 (1998).
3. Kristenson, M. et al. *Psychosom. Med.* **60**, 277–282 (1998).
4. Tunstall-Pedoe, H. et al. *Circulation* **90**, 583–612 (1994).
5. Jackson, R. et al. *Cardiovasc. Risk Factors* **7**, 43–54 (1997).
6. Weidner, G. & Collins, R. L. in *Attention and Avoidance* (ed. Krohne, H. W.) 241–265 (Hogrefe & Huber, Seattle, 1993).
7. Solomon, S. D. et al. *J. Appl. Soc. Psychol.* **17**, 1092–1112 (1987).
8. Kopp, M. *Psychosom. Med.* **60**, 103, abstr. (1998).

Missing the mark on misconduct

Sir—*Nature* devotes an editorial to the topic of “scientific misconduct”, takes a swipe at Congressman John Dingell and records that the US president's National Science and Technology Council has not been able to define the term in two-and-a-half years (*Nature* **394**, 815; 1998).

As a student of science policy, I submitted to the council for consideration (as I do the entire science community) the question: why is “scientific misconduct” confined only to communication by scientists to scientists in scientific journals? What is the evidence that any substantial harm has been done to the progress of science, or to society which foots the bill? By these statistically infrequent and scientifically insignificant examples we only prove that scientists have average human behavioural characteristics.

It has been my thesis that what the science community must give more attention to is the scientific misconduct that is vastly more significant to society and to the total ecology of science. I refer to the misconduct in irresponsible or misleading communications to the public. Alvin Weinberg, founding director of Oak Ridge National Laboratory and wise analyst of science policy, long ago wrote of the absolute necessity for means “to keep scientists honest and mechanisms for injecting more responsibility into the scientific debate, especially when it is

conducted outside the scientific forum", that is, in all dealings with the public.

Let us compare the two 'victim' populations. Each specialist science community is fully equipped and prepared to detect fraud (or honest error) in any paper. But the public is utterly defenceless against any exaggerated claim or hype, however egregious. Moreover, there may be enormous public consequences, such as the misdirection of billions of dollars resulting from such fraud or scientific misconduct because it influences the public.

It is also not possible for scientists to blame journalists for exaggerating their claims, unless they publicly disown them when they appear. I submit that all persons or bodies thinking about scientific misconduct should concern themselves first with the ethics of our behaviour when we deal with the public.

Rustum Roy

102 Materials Research Laboratory,
Pennsylvania State University, University Park,
Pennsylvania 16802-4801, USA

Question marks over genetic counselling

Sir— Perhaps the high tide of genetic determinism in all things biological is beginning to recede a little at last. I was relieved to read of the caution advised by the UK bioethics committee with regard to genetic screening as a means of predicting the susceptibility of individuals to mental disorders (*Nature* **395**, 309; 1998).

Even the most fervent supporters of the contentious idea that there is a significant genetic component to mental disorders would probably be prepared to admit that the correlation between the occurrence of a gene and that of a disorder in these cases is statistical. There are plenty of people with the gene, but not the disorder, and plenty of others with the disorder, but not the gene. This is also true of the much publicized genes 'for' heart disease and breast cancer, among others.

What does the individual do with the knowledge that they carry such genes and what advice can genetic counsellors give them? Don't smoke, don't drink too much, be careful about what you eat, take a little exercise, avoid stress and, you there with that gene, even more so? This is stretching the concept of genetic determinism beyond utility.

At a recent conference on the commercial potential of genomics I heard a representative of those who wish to offer such screening to all individuals on a commercial basis concede that their counsellors sometimes had difficulty in

communicating the importance of the information to the recipients. He appeared to think that this was a problem of education and perhaps he was right, though the problem may be his, not theirs. The recipients may have been sufficiently well educated to realize that statistics are properties of populations, not individuals, and, even if the assumptions about the genetic component of the disease were actually correct, the information they were receiving was absolutely useless to them.

Alan Akers

120 Avenue de Strasbourg,
67170 Brumath, France

Germany keen to reduce the nuclear threat

Sir— The German government is keenly interested in diminishing the dangers arising from the enormous stocks of plutonium no longer required for nuclear weapons and still held by certain states. It has taken an active part in all efforts to find ways of finally disposing of this material and is well acquainted with the ideas put forward by Frank N. von Hippel in your Commentary, "How to simplify the plutonium problem" (*Nature* **394**, 415–416; 1998).

I categorically reject the author's outrageous and totally spurious charge that Germany has "pursued the development of nuclear weapons under the cover of 'civilian' plutonium programmes". I refute his association of Germany with nuclear pariah states such as Iraq, North Korea or Pakistan and his insistence that "the danger is not past". Comments of this nature fly in the face of all Germany's declarations, commitments and treaty obligations since its accession to the Western European Union and the North Atlantic Treaty Organization as well as the Nuclear Non-Proliferation Treaty.

Moreover, they completely disregard decades of Euratom and International Atomic Energy Agency reports and safeguards inspections confirming the absolutely peaceful character of Germany's activities in nuclear research and power generation. These allegations are entirely devoid of foundation.

Martin Erdmann

Federal Foreign Office,
Auswärtiges Amt, Postfach 1148,
53001 Bonn, Germany

Frank N. von Hippel replies— I included Germany (and Sweden) in a long list of countries which I said "have all pursued the development of nuclear weapons under the cover of 'civilian' plutonium programmes". The following sentence began:

"Fortunately, internal political changes and external pressures have aborted most of these programmes..."

Germany and Sweden both abandoned their nuclear-weapons programmes before they signed the non-proliferation treaty of 1970, almost 30 years ago. I am sorry that the reader misunderstood me as impugning Germany's — or Sweden's — subsequent faithful adherence to that treaty. That was not my intention.

Similarly, my phrase that the "danger is not past" was not aimed at Germany.

Indeed, the next sentence discusses the danger that the spread of reprocessing in east Asia could exacerbate the danger of proliferation there.

Frank N. von Hippel

Center for Energy and Environmental Studies,
H-102 Engineering Quadrangle,
Princeton University,
Princeton, New Jersey 08544, USA

'No controversy' at CITES

Sir— I wish to comment on your article about the Convention on International Trade in Endangered Species (*Nature* **394**, 112; 1998). The CITES secretariat is not "one of the United Nations' most controversial secretariats", but rather quite the contrary. For 25 years, it has been a model of efficiency and qualified service.

Your statement that "two members of the CITES secretariat in Geneva have been dismissed" is incorrect. The staff in question opted for early retirement, and were offered compensation for their long years of service. The departure of these two professional staff members is highly regretted.

You refer to the role of these two individuals "in awarding permits to organizations that wanted to trade in plants and animals on the CITES list of banned species". But the CITES secretariat does not grant CITES permits; this is the role of the CITES management authorities of member states. Permits are not granted to "organizations". Nothing of this sort ever happened in the secretariat, either related to these two staff members or, to my knowledge, to any other member of the secretariat.

The statements in your article have damaged the reputation of two honourable and highly skilled professionals, who have left the secretariat to the regret of all the parties, and who are very much respected by all who have known them. They are also damaging to the CITES parties, and to the United Nations Environment Programme.

Victoria Lichtschein

Dirección de Fauna y Flora Silvestres,
Secretaría de Recursos Naturales y Desarrollo
Sustentable, Argentina

Strategies for cutting carbon

David G. Victor

What can be done to slow global warming? Huge new sources of carbon-free power may be needed. But other options also exist, and with so many uncertainties dogging predictions of technology and climate, choosing the best portfolio is hard.

The Earth is getting warmer, probably as a side-effect of human industry. The main culprit is carbon dioxide (CO₂), a by-product of burning fossil fuels for energy. On page 881 of this issue¹, Hoffert *et al.* examine one plausible scenario for future energy demand. They conclude that to stabilize the concentration of CO₂ in the atmosphere at twice pre-industrial levels will require a vast increase in the supply of carbon-free energy sources such as solar, wind and nuclear power.

In order to explore where policy-makers have leverage, Hoffert *et al.* express CO₂ emissions as the product of four variables: CO₂ released per unit of energy; energy consumed per unit of economic output; economic output per person; and the number of people.

There are good reasons not to meddle with the last two variables. Ideally, policy should limit CO₂ concentrations while doing minimum harm to the economy. Nor can politicians do much more to curb world population, which will probably stabilize at about 11 billion people². Many countries may even attempt to increase their numbers in order to reverse the strains on productivity and social security that result when populations shrink and grey.

That leaves policy-makers two options: 'decarbonize' the energy system by lowering the quantity of carbon emitted per unit of energy³, or improve energy efficiency (cut-

ting the primary energy needed per unit of economic output).

Both of these numbers have improved over the past 150 years. On average, worldwide, the energy consumed per unit of economic output has declined at about 1% per year⁴. Hoffert *et al.* show that, if that rate remains constant, then stabilizing atmospheric CO₂ at 550 parts per million will require about 15 terawatts (TW) of carbon-free power by 2050, and even more thereafter. For comparison, today the world consumes about 13.5 TW of power altogether, of which no more than 3.4 TW are carbon free (Fig. 1). Creating so much carbon-free power would require a decrease in the carbon emission per energy unit of about 1% per year. The historical rate has been a more leisurely 0.3% per year³. The gap,

conclude the authors, is huge.

This arithmetic is hardly controversial, but it leaves boulders unturned. Most importantly, Hoffert *et al.* omit economics from the analysis and so cannot explore the trade-offs that real policy must confront — such as the balance between carbon-free power and improved energy efficiency. For example, if policies to improve energy efficiency could accelerate the 1% annual decline in energy use per unit of economic output to 1.5%, then the carbon-free energy required would be cut in half (see Fig. 3 on page 883), and the urgency of investing in new power sources would wane. The energy system is full of fat: the useful energy output of many devices, such as illumination or locomotion, is only a few per cent of total energy input⁵. So large savings are quite possible.

Hoffert *et al.* may overstate the need for carbon-free power for another reason. Their baseline is the 1992 "a" scenario published by the Intergovernmental Panel on Climate Change, which assumes that coal, the most carbon-intensive fossil fuel, remains one of the main energy sources. So carbon emissions would be high, and stabilizing atmospheric CO₂ requires a lot of carbon-free power. But it may well be that even without a deliberate policy to slow global warming, energy production comes to be dominated by less carbonaceous natural gas, or by

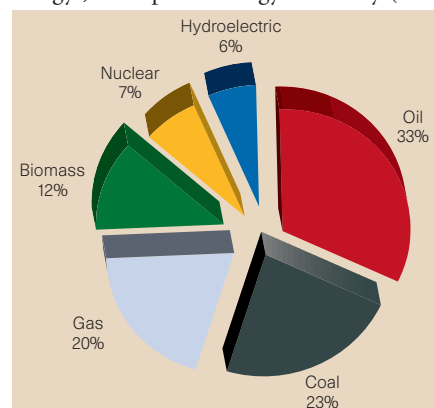


Figure 1 Global energy consumption in 1996, totalling 13.5 terawatt hours. About 25% is carbon free. Of the rest, coal releases 790 million tonnes of carbon per terawatt hour; oil 610; gas 470.



Figure 2 Three ways to curb carbon: nuclear fusion (top left), a carbon-free technology that is still not commercially viable; tree plantations (top right), which can remove carbon from the atmosphere or provide fuel wood with practically no net carbon emission; and energy-efficient light bulbs. Deciding on which are best depends on many factors, including expected future costs.



carbon-free nuclear and biomass fuels⁶, yielding much lower carbon emissions and less need to supply still more carbon-free power.

Politicians should also consider options for withdrawing carbon from the atmosphere by planting trees, improving soil management and disposing of CO₂ captured from the smokestacks of power plants before it is released to the atmosphere. That would lower the need for carbon-free power and is also likely to be a cost-effective option, especially during the next few decades, before radically new energy systems are available (Fig. 2).

Hoffert *et al.* are right to emphasize that to curb global warming will entail massive technological change. Research and development is needed to make new technologies viable, so it is worrisome that public energy-related R&D is declining in nearly every industrialized country^{7,8}. In the United States — the biggest spender on energy R&D — the total funding fell by 40% from 1985 to 1994, and worse was to come: gas and electricity companies cut basic research by two-thirds⁹ from 1995 to 1996, as restructuring and deregulation of energy markets led them to concentrate on short-term returns. Society as a whole prospers by the new concepts that emerge from basic research, but competitive companies have little incentive to make the necessary long-term investments. Nevertheless, efforts in some countries to reverse the trend are bearing fruit.

To slow global warming will require policies that penalize carbon emission and encourage higher investment in energy R&D. Those policies can be more effective if guided by economic and engineering analysis, but only if the models used to assess long-term global-warming costs and benefits include an improved representation of technological change¹⁰. Research is also needed to identify the best policies for directing

R&D. Hoffert *et al.* suggest something on the scale of the Manhattan Project or Apollo Program as models. But unlike those publicly funded crash programmes, an effective energy R&D programme must heed market conditions — after all, the market will dictate which technologies are eventually adopted.

And there is an even more fundamental question to be answered: what should be the objective of global-warming policy? Many studies assume that we should aim to stabilize atmospheric CO₂ at 550 parts per million, a convenient number that is about twice the concentration before the industrial revolution, and about 50% more than today's. But there is little solid evidence to justify that choice of number; nor is it clear whether any single number can express the damage that could result from global warming. For now, taking action on global warming is akin to buying insurance with an unknown premium against unknown hazards. Investing in new technology may be a good hedge against that uncertain future, but exactly what to do is still unclear. □

David G. Victor is at the Council on Foreign Relations, 58 East 68th Street, New York, New York 10021, USA.
e-mail: dgvector@cfr.org

1. Hoffert, M. I. *et al.* *Nature* **395**, 881–884 (1998).
2. Lutz, W., Sanderson, W. & Sherbov, S. *Nature* **387**, 803–805 (1997).
3. Nakićenović, N. *Daedalus* **125**, 95–112 (1996).
4. Manne, A. & Richels, R. *Buying Greenhouse Insurance: The Economic Costs of CO₂ Emission Limits* (MIT Press, Cambridge, 1992).
5. Nakićenović, N. *et al.* *Energy* **18**, 401–409 (1993).
6. Nakićenović, N., Grübler, A. & McDonald, A. (eds) *Global Energy Perspectives* (Cambridge Univ. Press & World Energy Council, 1998).
7. International Energy Agency, *IEA Energy Technology R&D Statistics: 1974–1995* (IEA, Paris, 1997).
8. Dooley, J. J. *Energy Policy* **26**, 547–555 (1998).
9. Morgan, M. G. & Tierney, S. F. *Issues in Science and Technology* **15**, 81–87 (1998).
10. Grübler, A. *Technology and Global Change* (Cambridge Univ. Press, 1998).

object, mostly from different angles^{2–5}.

Such is not the case with weakly electric fish. These nocturnal animals inhabit freshwater ponds and streams in South America and Africa. They produce weakly electric fields with an electric organ situated in the tail region and sense the distortions caused by the surrounding objects with their electroreceptors. Depending on its shape, location, electrical properties and distance, an object will alter the sensory feedback of a fish's electric organ discharges in a characteristic fashion, resulting in a specific 'electrical image' of the object on the fish's body surface. This image is then scanned by the array of electroreceptors in the fish's skin that transfer the information to the brain. The brain deciphers this code and determines the type and location of the object.

A problem in this context, however, is the ambiguity between object size and distance: a particular object might appear to be small because it is far away or simply because it is small. Are electric fish able to resolve this dilemma? And if so, how? In a series of behavioural discrimination experiments, von der Emde *et al.*¹ show that their fish can indeed determine the distance of objects independently of the objects' size, shape or constituent material. Surprisingly, however, when the authors measured the electrical images produced by various cubes, pyramids and spheres placed in an animal's electric field, they found no single physical parameter that could have provided the fish with an unambiguous cue. But after further data analysis, they identified one particular combination of parameters that was unequivocally correlated with object distance: the ratio of the 'slope' of the electrical image (that is, how 'fuzzy' it is at the edges) over its maximum amplitude.

To make this intuitively a bit more comprehensible, compare this combination of electrosensory parameters with visual cues. Unlike visual objects, electrical objects create larger images the further away they are. So, as shown in Fig. 2, an electrical image is in this regard better compared to the shadow an object casts on a screen: the size of the shadow increases with distance and the edges of the shadow become fuzzier. This results in a lower ratio of maximum darkness of the image over maximum slope of intensity change at the image's edge; it could, thought von der Emde *et al.*, be the way in which weakly electrical fish sense distance.

To test their hypothesis, the authors applied behavioural experimentalism at its best — they cheated the experimental animal. This is the highlight of the paper. The authors took advantage of an 'electrical illusion': their slope/amplitude measurements revealed that spheres always yielded slightly lower ratios than any other object shape, because the round contour of a sphere creates a less intense and more

Perception

Measuring distance in two dimensions

Walter Metzner

Depth perception is a tricky task. A three-dimensional environment is first mapped onto a two-dimensional array of receptor cells (such as the photoreceptors in eyes; auditory receptors in the inner ear; mechanoreceptors in the lateral-line system of fish and aquatic amphibians; or, in some fish, also electroreceptors). Animals then have to reconstruct a three-dimensional image of their world in the brain, and they have come up with many ways to do so.

The work by von der Emde *et al.* (page

890 of this issue¹) shows us yet another way that nature has found to achieve this end. The authors show that weakly electric fish (*Gnathonemus petersii*; Fig. 1) can measure the distance of stationary objects unambiguously by instantaneously analysing the electric image of objects with a single array of electroreceptors embedded in their skin. Other animals manage the same job equally well only by using at least a pair of sense organs, such as the two eyes; or, if they use only a single sense organ, they take two or more snapshots of the same



Figure 1 The weakly electric fish, *Gnathonemus petersii*. The pores of the electroreceptor organs are visible as small white dots on the fish's skin.

uniform electrical image than the sharp edges of, for instance, a cube (the electric flux at sharp points is higher). Hence, when a fish is trained to choose between a sphere and, say, a cube that are both at the same distance, the sphere should appear to be further away. Conversely, the fish should not be able to discriminate between the two objects when the sphere is actually further away. This difference in the distance of the two objects should correspond to the difference extracted from their slope/amplitude ratios. Indeed, this is in essence what von der Emde *et al.* found.

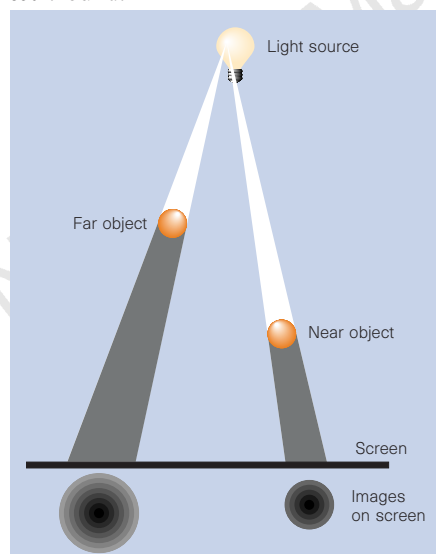


Figure 2 Electric image formation, exemplified by the formation of shadows on a screen. The further away an object, the larger the image and the fuzzier its edges. In finding out that weakly electric fish can judge the distance of objects, von der Emde *et al.*¹ observe that the ratio of the slope of the increased darkness of the image (its 'fuzziness'), over its maximum darkness, is the parameter combination that characterizes the fishes' performance. The fuzziness of the image alone was not an unambiguous indicator.

An alternative idea⁶, based on electric-field measurements and computer modelling, about how electric fish could perceive depth was proposed a couple of years ago. This mechanism would use the ratio of how far the electric image spreads over the electroreceptor array (its 'width'), divided by the amplitude of the electric image. This theory, however, fails to explain the fish's discrimination performance under certain experimental conditions⁷.

When compared with the way in which most other animals determine the distance of objects, von der Emde and colleagues' proposed mechanism is special also because it does not seem to require the brain to receive sensory input from the two arrays of electroreceptors on the left and right sides of the body (this will have to be confirmed, however, for instance with denervation experiments). Particularly in the visual world, the brain usually computes depth by analysing disparities between two (or more) eyes; anyone who has had to wear an eye-

patch for a while can attest to the trouble one can have when reaching for something.

Animals that can accurately judge the distance of stationary objects with one eye alone also use radically different techniques. Chameleons and frogs, for example, use accommodation mechanisms (that is, a proprioceptive feedback signal), and measure the state of focus of the eye lens to gauge distance^{2,8}. Mantids and locusts use peering movements generated by a sideward-swaying of head or body to induce displacements of the image on the retina⁴. This motion parallax is probably the most important method available to insects to judge distance.

In the auditory world, spectral and/or intensity cues from only one or the other ear are commonly used to extract the distance of a sound source. Spectral cues are also used by surface-feeding fish that detect water perturbations with their mechanosensory system to determine the distance of prey⁹. Finally, bats, dolphins and two groups of birds use the time domain to reconstruct the third spatial dimension, and listen to the delay between call emission and the returning echo signal^{3,10}. Electric fish have now taught us that distance can also be measured instantaneously, never leaving two-dimensional space. □

Walter Metzner is in the Department of Biology, University of California at Riverside, Riverside, California 92521-0427, USA.

e-mail: walter.metzner@ucr.edu

1. von der Emde, G., Schwarz, S., Gomez, L., Budelli, R. & Grant, K. *Nature* **395**, 890–894 (1998).
2. Collett, T. S. & Harkness, L. I. K. in *Analysis of Visual Behavior* (eds Ingle, D. I., Goodale, M. A. & Mansfield, R. J. W.) 111–176 (MIT Press, Boston, MA, 1982).
3. Griffin, D. R. *Listening in the Dark* (Cornell Univ. Press, Ithaca, NY, 1986).
4. Horridge, G. A. *Proc. R. Soc. Lond. B* **229**, 13–27 (1986).
5. Rossell, S. *Proc. Natl Acad. Sci. USA* **93**, 13229–13232 (1996).
6. Rasnow, B. J. *Comp. Physiol. A* **178**, 397–411 (1996).
7. Schwarz, S. Thesis, Univ. Bonn (1997).
8. Ott, M. & Schaeffel, F. *Nature* **374**, 692–694 (1995).
9. Bleckmann, H., Tittel, G. & Bluebaum-Gronau, E. in *The Mechanosensory Lateral Line* (eds Coombs, S., Goerner, P. &

Materials science

Crystal cages for clean coolers

Laszlo Mihaly

You may remember a demonstration experiment in some introductory physics class, when the instructor was furiously shaking the end of a rope that was hung across the lecture room, and waves were propagating from one corner of the room to the other. On a microscopic level, similar processes carry heat in solids. In an emerging area of materials research the name of the game is to suppress these processes, so that more of the heat is made to ride with an electrical current. How can we make new materials that are disordered and dirty with respect to the lattice vibrations

that carry heat only, but look nice and clean for the electrons that do all the useful work? Keppens *et al.*, on page 876 of this issue¹, show us one way to do it.

There are two main thermoelectric phenomena: the Seebeck effect, in which a temperature difference along a conductor produces a voltage; and the Peltier effect, where a current causes a temperature difference. We have known of these effects for more than 100 years. The fundamental laws of irreversible thermodynamics, discovered by Lars Onsager in the late 1920s, provide a mathematically simple yet physically deep

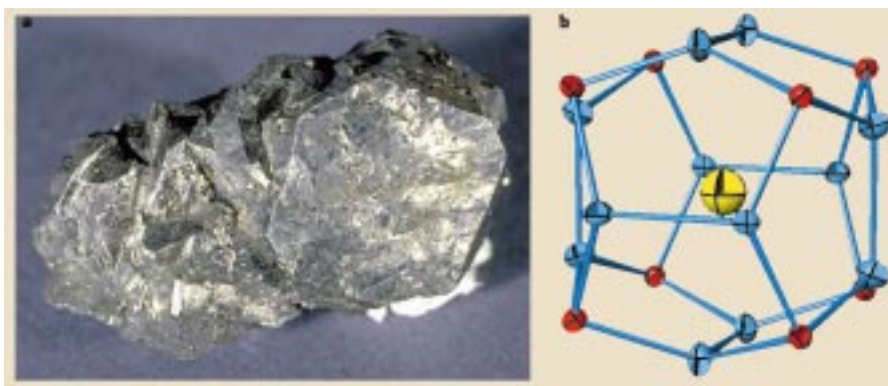


Figure 1 A piece of skutterudite (a), and the atomic arrangement of a filled skutterudite (b). Rare-earth atoms (yellow) are thought to rattle in the cages formed by iron or cobalt (blue) and antimony or arsenic (red), disrupting lattice vibrations and so reducing thermal conductivity. That could make these compounds ideal candidates for thermoelectric applications. These local vibrations have now been demonstrated, but they are unexpectedly complicated.

relationship between the corresponding materials parameters: $S = P/T$, where T is the absolute temperature, and the Seebeck coefficient S and the Peltier coefficient P are defined as the ratio of the voltage to the temperature difference and the heat flow to the electric current, respectively.

Thermoelectricity is used in all kinds of devices: in thermometers, in compact nuclear energy sources on satellites, for corrosion protection of pipelines, to transfer heat away from computer chips, and as a cooling element in portable refrigerators, night vision scopes and infrared sensors. Because of the lack of chemicals and moving

parts, these devices tend to be reliable and simple. But their usefulness is so limited by poor efficiency that the widespread application of thermoelectricity has been called "a breakthrough that never came"². During the past few years, however, progress in the theory of electrical and heat transport, together with the discovery of new materials, has renewed interest in this field²⁻⁶.

For good efficiency, a large thermoelectric effect, by itself, is not enough. Undoped semiconductors, for example, have large S values, but the low electrical conductivity σ makes them unable to carry high currents and supply much power. Similarly, a material

of large P can pump the heat out of a refrigerator, but a high thermal conductivity κ in the same material will let too much heat back in. The real usefulness of a compound is best measured³ by the 'coefficient of merit', $Z = \sigma S^2 / \kappa$. Notice that σ and S are entirely due to the motion of electrons, but the heat is conducted by both the electrons and the lattice vibrations.

The efficiency of a device depends^{2,3} on the dimensionless product ZT . There is no strict theoretical upper limit here, and a ZT of about 3 would make a Peltier refrigerator competitive with the traditional, compressor-based systems. For pure elements ZT is much greater than 1, and the compounds in use today have a ZT of about 1. Can we improve the coefficient of merit by clever design? The key is to minimize the lattice contribution to thermal conductivity. Lattice vibrations cannot be eliminated, but they can be localized by introducing lattice defects, up to a limit set by amorphous materials⁷. (In truly amorphous materials, the electrical conductivity and thermoelectric effect are too small to be useful.) As far as the electrons are concerned, high mobility and near-perfect crystal structure are preferred.

A class of compounds called skutterudites may be able to provide these contradictory properties. Skutterudites, such as CoAs_3 and IrSb_3 , have a crystal structure with unusually large, empty cavities. A 'filled skutterudite' is obtained when lanthanide atoms are introduced into some of these gaps

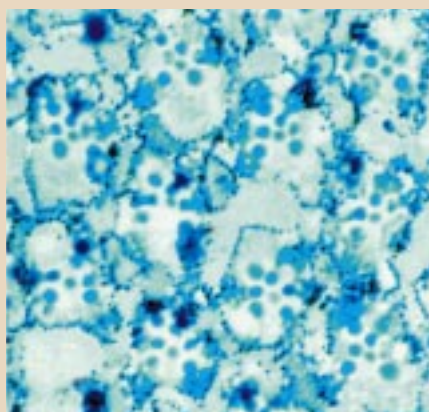
Neurodegeneration

A fly's eye view of Huntington's disease

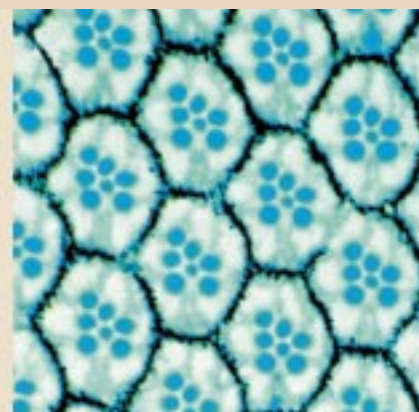
The picture on the left depicts neurodegeneration in the fruitfly *Drosophila melanogaster*, caused by expression of a human protein in its compound eye. Mutations in this protein are responsible for the inherited neurodegenerative disorder Huntington's disease, and these flies should be a useful animal model for the human condition.

In fact, there is already a mouse model for Huntington's disease, so why do we need another? Writing in *Neuron* (21, 633-642; 1998), George R. Jackson *et al.* explain that, whereas mutant mice do not show much in the way of neuronal cell death, the flies show many of the same symptoms as humans.

Huntington's disease is caused by long tracts of the amino acid glutamine in a protein called huntingtin. In healthy people this protein contains a run of 37 or fewer glutamine residues. But in patients with Huntington's disease that number can exceed 150. Protein fragments containing the glutamine tracts are toxic to nerve cells, killing them through the formation of nuclear aggregates.



Might long glutamine tracts also cause neurodegeneration in *Drosophila*? To test this, Jackson and colleagues expressed fragments of human huntingtin complementary DNA containing 2, 75 or 120 consecutive glutamines in adult fly eyes. *Drosophila* eyes consist of an array of cell clusters called ommatidia (pictured right), each containing eight photoreceptor neurons. The authors found that expression of the 120-glutamine tract leads to rapid (within ten days) degeneration of



the photoreceptor cells (pictured left). The 75-glutamine tract caused much less severe disruption after ten days, consistent with the human condition — the longer the repeat, the sooner the disease takes hold.

Jackson *et al.* also show that death is not blocked by expression of an anti-apoptotic protein, p35, suggesting that the death pathway is distinct from those already known in flies. But what could this pathway be? Perhaps the *Drosophila* model will help to answer these questions. Alison Mitchell

(Fig. 1). But as the cavity is larger than the atom, the lanthanide is not confined to a well-defined position. According to Glen Slack and V. G. Tsoukala^{3,4}, as it rattles around in its cage it may destroy the large-scale coherence of the lattice vibrations that carry heat, localizing the vibrations instead. Yet because the conduction electrons do not overlap much with the lanthanide, electrical transport processes are barely disturbed.

Keppens *et al.*¹ have now proved the existence of such localized rattling in several experiments. Specific-heat measurements, ultrasonic spectroscopy and inelastic neutron scattering confirm the existence of a low-energy resonance that can only be ascribed to the rattling of the lanthanide atom in its cage. The full story, however, proves to be more complex than expected. Instead of seeing just one resonance, the authors find two, the second being broader and at higher energy. Furthermore, the old description of this system in terms of Einstein oscillators — a relatively simple mathematical model — is not satisfactory. Quantum mechanical tunnelling between atomic configurations, already seen in the physics of amorphous materials, seems to play an important part here as well.

A complete interpretation of these experiments will certainly tax existing theories of lattice vibrations. The results of Keppens *et al.* have no direct implications for the design of better thermoelectric devices, but as our understanding of the basic physics of filled skutterudites deepens, the hunt for a higher coefficient of merit should intensify. Many laboratories tooled for materials research on high- T_c superconductors are ready to expand into the re-emerging field of thermoelectric compounds. So perhaps in the not-too-distant future we will be able to have an air-conditioned car without making the worrisome hole in the ozone layer⁸ even larger. □

László Mihály is in the Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3800, USA.
e-mail: lmihaly@ccmail.sunysb.edu

1. Keppens, V. *et al.* *Nature* **395**, 876–878 (1998).
2. Mahan, G. D. in *Solid State Physics* Vol. 51 (eds Ehrenreich, H. & Spaepen, F.) 82–158 (Academic, San Diego, 1998).
3. Slack, G. A. in *CRC Handbook of Thermoelectrics* (ed. Rowe, D. M.) 407–440 (CRC Press, Boca Raton, 1995).
4. Slack, G. A. & Tsoukala, V. G. *J. Appl. Phys.* **76**, 1665–1671 (1994).
5. Nolas, G. S., Cohn, J. L. & Slack, G. A. *Phys. Rev. B* **58**, 164–170 (1998).
6. Sales, B. C. *et al.* *Phys. Rev. B* **56**, 15081–15089 (1997).
7. Cahill, D. G., Watson, S. K. & Pohl, R. O. *Phys. Rev. B* **46**, 6131–6140 (1992).
8. <http://toms.gsfc.nasa.gov>

Human genetics

A map for cyberspace

Peter Little

Twenty-five years ago we could only dream of identifying human genes that, when perturbed in some way, cause disease. Today it is a routine but difficult process. Newly published work¹, or rather the associated website², is going to

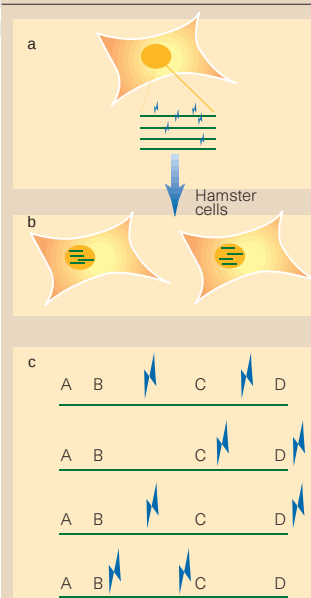
make the routine less difficult, for it gives the positions of more than 30,000 genes along the human DNA molecule, starting with the genes at one end and stopping at the other. At a stroke this doubles the number of genes in known positions.

This remarkable feat was achieved by combining something old and something new. The old is an idea first explored in 1975 (ref. 3) and the new is the information streaming from the cutting edge of the human genome project. The authors have used radiation hybrids, first described in the 1975 paper, to assign locations to human genes along the strand of DNA. The details are given in the box below, “Radiation hybrids and gene order”, but there is no need to understand the technicalities. All that needs to be grasped is that gene sequences have been taken from the publicly available DNA databases (particularly those of the complementary DNA sequences such as dbEST; ref. 4) and, using the methods detailed in the box, genes have been grouped together on the DNA strands.

This has enabled the researchers to establish a linear order of 30,181 genes in all. The location is not quite perfect, in that the genes are not organized 1,2,3,4 ... 30,181; rather, they are ordered in clusters, 1–4 coming before 5–7, which come before 8–12 and so on. Because of the statistical methods used in mapping radiation hybrids, there is also a possibility that some orders are wrong. Fortunately, this imprecision is overcome by the types of experimental analyses geneticists routinely carry out.

What’s the point of all of this? Why not wait for the full, continuous DNA sequence from the genome project due in the first few years of the new millennium? The answer is that knowing where a gene is located in our DNA is the key to identifying genes that cause disease. The standard approach is to use large-scale DNA analyses to identify roughly where in our DNA the gene causing a genetic disorder is located. Each gene in the vicinity then has to be identified, which

Radiation hybrids and gene order



The figure shows how 30,181 human genes have been put in order, in clusters, along the human DNA strand¹².

a, Human cells are treated with X-rays, breaking the DNA (green) into fragments of 2.5–25 million base pairs. This kills the cells, but before they disintegrate they are fused with recipient hamster cells which can continue to grow. b, Each recipient hamster cell takes up 15–25% of the pieces of human DNA. The recipients are viable and called radiation hybrids. A total of 176 single cells, chosen to fit into two 96-well plates (a robotic standard),

are isolated and separately grown in bulk – each cell line has a different complement of fragments but each has 15–25% of human DNA. c, DNA analysis: all of the 176 radiation hybrids are tested, using the polymerase chain reaction, for the genes A–D to see if the original cell contained their DNA. Because only 15–25% of DNA was taken up by each hybrid, only a subset will contain A, B, C or D. If A and D are a long way apart in normal, untreated DNA, then the chances are high that they will be on separate X-ray-induced

fragments and will not often be found in the same hybrid cell. In contrast, if A and B are close together, they will always occur on the same fragment and therefore in the same radiation hybrids. So we can say A, B, C, D is the likely gene order based upon the frequency of co-occurrence of A+B, A+B+C and A+B+C+D. Note that two genes, A and D, even though far apart, could still be found, by chance, in the same cell. This means that the new map¹² is based on a statistically, not absolutely, defined order of genes.

P.L.

is colossally laborious; having done this, the hapless researcher then has to analyse every base pair in every gene to find out if it is altered in people with the disorder. If a gene can be identified that is always changed in affected (or carrier) individuals, but never in unaffected individuals, this is formal proof that the gene contributes to the disorder. Finding the culprit gene in any area is enormously difficult and heartbreaking — all except one will usually turn out to be merely a disease-associated ‘candidate’, not a causative, gene.

Hence the significance of the new map^{1,2}: we now know the location of about half of our genes, making the task of gene-finding much easier. Instead of laborious, ‘wet’ research work, we can use a computer to do part of the job and carry out electronic gene-hunting in the gene-location database². Of course, half our genes is better than none, but what about the other half? Fortunately, the authors have made it possible for you to do your own analysis of the radiation hybrid DNAs and to use two web-based analysis programs^{5,6} to locate your own favoured gene — an enormously powerful tool that is a new feature of this map compared with some previous ones.

Virtually all of the common single-gene disorders have yielded their causative gene, however, and the rare ones are just that — rare. Why spend the money on this latest exercise in genetic cartography? One reason

is because the really common genetic disorders, such as diabetes in its various forms, some cancers, asthma, migraine and coronary heart diseases, do not result from changes in a single gene but from changes in several genes. This makes the task of locating the genes much harder and, for technical reasons, the location much less precise. This in turn implies that researchers have to search a larger region and therefore sequence even more genes. Making this process more efficient will be a great help in understanding these common conditions.

A second reason is that the results of gene-mapping act as signposts in human DNA. We now have a far better framework for positioning the huge mass of base pairs that are coming out of the dozen or so genome-sequencing centres, and that are inexorably, and excitingly, accumulating towards the total of 3,000 million. Genes and computers are going to be an integral influence on our lives as scientists, patients, parents and citizens for many generations to come. □

Peter Little is in the Department of Biochemistry, Imperial College, Prince Consort Road, London SW7 2AZ, UK.

e-mail: p.little@ic.ac.uk

1. Deloukas, P. *et al.* *Science* **282**, 744–746 (1998).
2. <http://www.ncbi.nlm.nih.gov/genemap98/>
3. Goss, S. J. & Harris, H. *Nature* **255**, 680–684 (1975).
4. <http://www.ncbi.nlm.nih.gov/dbEST/index.html>
5. http://www-shgc.stanford.edu/RH/rhserver_form2.html
6. <http://www.sanger.ac.uk/RHserver/RHserver.shtml>

Galaxies

New members of the Local Group

Sidney van den Bergh

Three nearby galaxies have been discovered, by Armandroff *et al.*¹ and by Karachentsev and Karachentseva². All three are faint dwarf galaxies that lie within the Local Group³, the small cluster of galaxies dominated by three spiral systems — Andromeda (M31), Triangulum (M33) and our Milky Way. The discovery confirms that these dim dwarf galaxies are common in the Universe, but also raises questions about

how many galaxies exist that are even less luminous.

The newly discovered galaxies are ‘dwarf spheroidals’, much smaller than the more familiar spiral galaxies and large ellipticals. They also differ from other dwarf galaxies — dwarf irregulars — in that they are no longer forming stars. One of the two main sub-clusters of the Local Group, centred on the Milky Way system, is known to contain eight dwarf

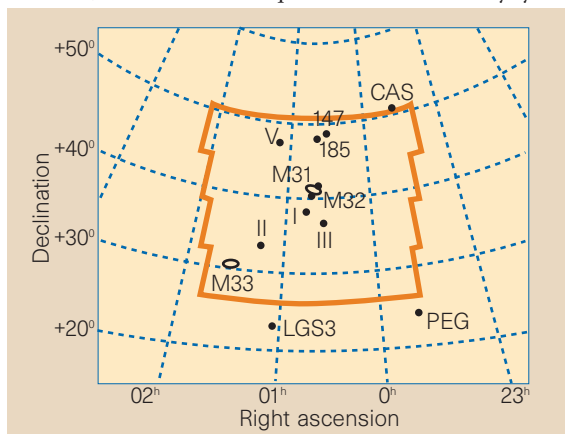


Figure 1 **Companions of Andromeda (M31).** The recently discovered dwarf spheroidal galaxies And V, Cassiopeia and Pegasus are even fainter than And I, And II and And III, which I discovered in 1971 in a survey of the outlined area⁴.



100 YEARS AGO

The following interesting announcement appears on a page in the catalogue of Messrs. Johnson, Matthey, and Co., Hatton Garden, London — “In furtherance of scientific research, Professors and recognised scientific investigators will with pleasure be supplied with metals of the platinum group, in moderate quantities, and for periods to be arranged, free of charge, on condition that the precious metals are ultimately returned (in any form), and that the results of the investigations are furnished.”

The coasts of Japan are particularly liable to incursions from spring tides, of which one occurring on June 15, 1896, in the course of eighteen minutes swept away 9381 houses and 6930 boats, killing 21,909 people and wounding 4398. To minimise the damage done to life and property by such inroads, protective forests have been planted at various places along the littoral. ... The action of these forests is three-fold: they check the force of the tidal wave; they delay its advance, giving more time for saving the lives of inhabitants living behind the forest; and, lastly, they prevent houses and property from being washed away into the sea.

From *Nature* 27 October 1898.

50 YEARS AGO

The second season's work of the British-Kenya Miocene Expedition in the Kavirondo region of Lake Victoria has culminated in one of the most important discoveries yet made there. Dr. L. S. B. Leakey, the field director of the Expedition, has announced the finding on Rusinga Island on October 2 of the greater part of a skull of one of the species of Miocene apes belonging to the genus *Proconsul*, probably *Pr. africanus* (Hopwood). ... [T]his new discovery for the first time provides information regarding the whole of the facial skeleton and much of the brain case. ... Mrs. Leakey was actually the first to see some small fragments of the skull, where they had been washed out on the slope of one of the gullies which were being explored. She directed the attention of her husband who, cutting back into the beds, brought to light this most important, and indeed unique, fossil.

From *Nature* 30 October 1948.

spheroidal galaxies. So it was surprising that only three such objects were known in the second subgroup, around Andromeda — especially as Andromeda is more luminous than the Milky Way.

Looking in the Andromeda region, two groups of astronomers^{1,2} used a digitized version of the Palomar Sky Survey to search for dwarfs that were too faint to be seen on the original photographic plates. Two of the newly discovered dwarf spheroidals are also outside the region originally surveyed⁴: the Cassiopeia and Pegasus systems both lie more than 200 kiloparsecs from M31 (Fig. 1). This means that they may not be satellites of Andromeda, but instead free-floating members of the Andromeda subgroup.

So the new discoveries reduce the apparent difference between the two local subgroups, and the discrepancy may be further reduced after some faint suspected dwarf galaxies in the direction of the Andromeda subgroup are studied in more detail later this year. Furthermore, measurements of the velocities of faint members of the Local Group should give us constraints on the dynamics and total masses of the two subgroups.

This is of more than local relevance. Inspection of the Palomar Sky Survey shows that a large fraction of all galaxies occur in small clusters that resemble the Local Group, so we can use our neighbourhood as a sample of the Universe as a whole. Of the 35 presently known⁵ Local Group members, 16 are dwarf spheroidals and two are transitional objects somewhere in between spheroidal and irregular. This implies that half of all of the galaxies in the Universe may be dim dwarf spheroidals.

Could the fraction of dwarf spheroidals be much larger? Not according to a survey by Irwin⁶, which appears to show that few additional faint dwarf galaxies remain to be discovered in the Milky Way subgroup of the Local Group — there appears to be a cut-off at a certain luminosity. This conclusion might be tested by the study of new dwarf galaxy suspects in the Andromeda subgroup.

If this cut-off is real, it may correspond to a lower limit to galaxy mass. The velocity dispersions of individual stars in dwarf spheroidals⁷ have so far revealed no objects in which the total mass is less than ten million times that of our Sun. Is this a real lower limit to the mass of the dark matter in galaxies? Or it is due to the fact that all gas can escape rapidly from such low-mass galaxies before it forms many observable stars? Only future observations of even fainter dwarf galaxies will be able to answer these questions. □

Sidney van den Bergh is at the Dominion Astrophysical Observatory, 5071 West Saanich Road, Victoria, British Columbia V8X 4M6, Canada.
e-mail: vandenbergh@dao.nrc.ca

1. Armandroff, T. E., Davies, J. E. & Jacoby, G. H. preprint astro-ph/9807232 on xxx.lanl.gov
2. Karachentsev, I. & Karachentseva, V. *Dwarf Tales* No. 3, 1–2 (1998).
3. Hubble, E. *The Realm of the Nebulae* (Yale Univ. Press, New Haven, CT, 1936).

4. van den Bergh, S. *Astrophys. J. Lett.* **171**, L31–L33 (1972).
5. van den Bergh, S. *The Galaxies of the Local Group* (Cambridge Univ. Press, in the press).
6. Irwin, M. J. in *Dwarf Galaxies* (eds Meylan, G. & Prugniel, P.) 27–36 (Eur. Southern Obs., Garching, 1994).
7. Mateo, M. *Annu. Rev. Astron. Astrophys.* **36**, 435–506 (1998).

Neurobiology

In the mind's eye of the beholder

Robert Shapley

Where is colour — in the world around us or in our minds? Although we view colour as something objective in the world, students of visual perception have known for a long time that to perceive colour we need to make elaborate neural computations. The brain constructs a colour signal to recover, as well as it can, the true reflective properties of a given surface, independent of illumination or the surrounding colours. On page 896 of this issue, Cottaris and De Valois¹ report that they have used modern neurophysiological techniques to study how neural circuits in the primary visual cortex (V1) contribute to the mental process of constructing colour in the brain.

The primary visual cortex does much more sophisticated image processing than we ever imagined. V1 is the first stage in the

cerebral cortex to analyse visual signals from the retina. These signals are relayed to V1 by the lateral geniculate nucleus (LGN) of the thalamus. In humans, V1 is located just under the occipital bone in the back of the head, and it provides the adjacent visual cortical areas with all that they need to know about the visual world. It used to be thought that V1 acts simply, like a cortical retina, mapping the visual scene point for point and doing relatively benign operations (such as spatial filtering) on the representation of visual signals. But more recent research has shown that visual cortical cells may be understood better as dynamical systems rather than static receptive fields. Indeed, cortico-cortical interactions are thought to be important in the response of the primary cortex to visual scenes^{2–4}.

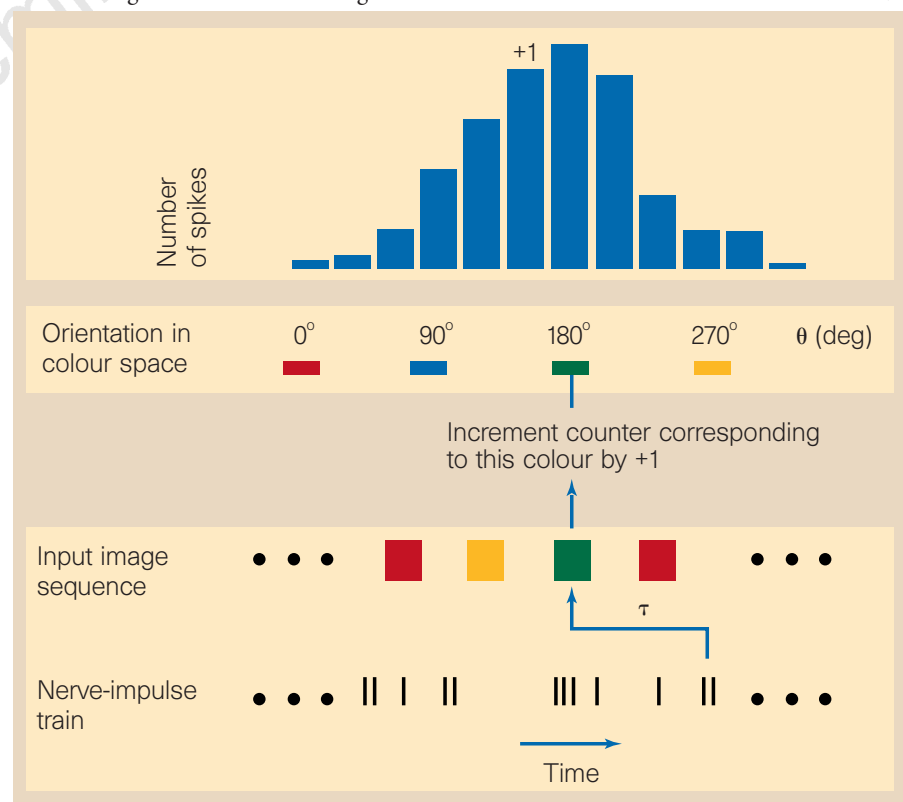


Figure 1 The 'reverse-correlation' approach to studying colour processing in cortical neurons used by Cottaris and De Valois¹. A sequence of coloured images is shown to a monkey and the action potentials from neurons are recorded. At delay time τ the coloured image preceding each spike is found and the counters in the histograms are increased once for each spike preceded by that colour. The resulting histogram is the distribution across colour space of the response for τ . This correlation is done for all values of delay to construct a response surface in the dimensions of colour-space angle and time delay.

Cottaris and De Valois¹ now provide an example of the 'reverse correlation' approach to studying the function of the cortex, similar to that used by Ringach *et al.*⁴, who explored cortical responses to orientation. Cottaris and De Valois used a sequence of rectangles as stimuli, the colours of which varied in time as if the sequence were random (Fig. 1). They recorded action potentials from single neurons in V1 of anaesthetized macaque monkeys, which seem to have eyes and V1 areas similar to our own. They found that neurons in V1 fire nerve impulses that correlate with the coloured pattern sequence. Moreover, by cross-correlating the neural spike trains with the colour sequence, they discovered which colours are likely to cause the neurons to fire.

The authors found that the representation of colour is reorganized in V1 neurons. The neural pathways from the LGN to cortex transmit colour signals along so-called opponent colour channels — classes of LGN neurons with similar colour properties. The name 'opponent' comes from the fact that the colour signal is the difference between signals from the three different types of cone photoreceptor. Most LGN neurons convey signals that are proportional to the difference between activity in the long-wavelength (L) and mid-wavelength (M) cones, and are called L – M cells. LGN input also comes from a sparser pathway, which carries signals proportional to the difference between activity in the short-wavelength (S) cones and the sum of activities of L and M cones. These are the S – (L + M) cells. Cottaris and De Valois now suggest that the influence of the S – (L + M) signal grows strongly within V1, relative to its representation in the input from the LGN to the cortex. The time course of its growth indicates that the greater influence of this colour signal in V1 is due to cortico-cortical coupling.

The authors also show that the response of cortical cells to colour is a rich, dynamic process, so it is important to study the colour-coding properties of cortical cells as their responses evolve. Besides the slow evolution of the S – (L + M) signal in V1, they also find, rarely, neurons whose colour preference slides smoothly from one region of colour space to another as a continuous function of time. Ringach *et al.*⁴ reported analogous V1 neurons whose preferred orientation varied continuously with time. The existence of such neurons — even if they are rare — is important because it suggests that cortical responses are generated by the summation, by the cortical neuron, of excitatory and inhibitory influences with different time courses. Models of orientation tuning^{5,6} that postulate the importance of lateral cortical connections are consistent with such dynamic shifts in stimulus preference in orientation, and it now seems that similar interactions occur in the domain of colour.

It is appropriate in this cinematic age that visual neurophysiologists should discover the usefulness of cinematic stimuli for studying the functional connectivity and the visual function of the cerebral cortex. To the cortex life is a movie, and we may be able to find out quite a lot about cortical function by studying how it responds in the screening room. □

Robert Shapley is at the Center for Neural Science, New York University, 4 Washington Place, New York, New York 10003, USA.
e-mail: shapley@cns.nyu.edu

1. Cottaris, N. P. & De Valois, R. L. *Nature* **395**, 896–900 (1998).
2. Pei, X. *et al.* *J. Neurosci.* **14**, 7130–7140 (1994).
3. Frégnac, Y. *et al.* *J. Physiol. (Paris)* **90**, 189–197 (1996).
4. Ringach, D. *et al.* *Nature* **387**, 281–284 (1997).
5. Ben-Yishai, R., Bar-Or, R. & Sompolinsky, H. *Proc. Natl Acad. Sci. USA* **92**, 3844–3848 (1995).
6. Somers, D. C., Nelson, S. B. & Sur, M. *J. Neurosci.* **15**, 5448–5465 (1995).

chromosomes during mitosis, and argued that it might help to maintain the structure and elasticity of these chromosomes.

Over 90% of titin is repeating amino-acid modules that are characteristically found in immunoglobulins and fibronectin. But it also contains a protein serine/threonine kinase domain near its carboxy terminus². This kinase domain is related to the myosin light-chain kinase family of calcium/calmodulin-binding proteins³, most of which show remarkable substrate specificity and phosphorylate a single residue in the regulatory light chains of myosin. In the absence of calcium/calmodulin, these kinases are inactivated by a protein segment occluding the substrate-binding region. This 'intrasteric autoinhibition' is relieved by binding of calcium/calmodulin, and the enzyme is activated. If calmodulin is removed through a decrease in intracellular calcium, however, autoinhibition is re-established. A structural basis for this autoinhibitory mechanism was first worked out by Kobe *et al.*⁶ two years ago. They solved the three-dimensional structure of twitchin kinase, the homologue of titin kinase in the nematode *Caenorhabditis elegans*. Surprisingly, regulation of titin kinase seems to be more similar to that of other calmodulin kinases than it is to twitchin kinase.

Calmodulin-dependent kinases such as calcium/calmodulin-dependent protein kinase I (CaMKI) are activated by a dual mechanism: calcium/calmodulin binding relieves autoinhibition and exposes a threonine residue in the 'activation loop' of the kinase, and this threonine is then phosphorylated by an upstream protein kinase⁷. Phosphorylation of this threonine in CaMKI, as well as similarly positioned residues in other protein kinases, redirects the activation-loop segment, allowing the protein substrate to interact with another residue in the activation loop — a strictly conserved aspartic acid that participates in catalysis. Indeed, phosphorylation occurs in all protein serine/threonine kinases with an invariant arginine (R) residue immediately preceding the catalytic aspartic acid (D). Hence, they have been collectively called RD kinases⁸.

The structure of the titin-kinase domain now solved by Mayans *et al.*¹ reveals a form of autoinhibition that has similarities to CaMKI, but also some differences. The similarities include the need for calcium/calmodulin to allow ATP binding and expose a single amino acid to an upstream protein kinase⁹. But the titin kinase domain is different in that it is phosphorylated on a tyrosine (rather than a threonine) residue. Moreover, this residue is found in the pocket that accommodates the P+1 position of the substrate — that is, in the 'P+1 loop' of the enzyme, rather than in the activation loop.

Studies with other kinases show that the

Muscle proteins

The clash in titin

Anthony R. Means

According to Greek mythology, the Universe was formed and controlled by Titans — godlike giants who created order from chaos. In the same way, the most gigantic of all proteins, titin, brings order to sarcomeres (the contractile units of striated muscle) and organizes the machinery for condensation of chromosomes in dividing cells. On page 863 of this issue, the article by Mayans *et al.*¹ gives us a first look into the face of titin, specifically at the protein-kinase catalytic domain. Their results provide a surprising new perspective on the structure and function of a kinase, and should help us to work out how titin kinase is regulated.

Titin, also known as connectin, is composed of over 27,000 amino acids², and each molecule spans half the distance of a sarcomere (~1.2 µm). Titin is tethered to the Z-disk — a dense line that separates one sarcomere from the next — by its amino terminus, and to the centre of the sarcomere (the M-line) by its carboxy terminus. Thus, titin seems to be a scaffold upon which sarcomeres are assembled, and it may also be responsible for the elasticity of striated muscle³. Earlier this year, Machado *et al.*⁴ showed that titin is recognized by the self-reactive antibodies produced in patients with the autoimmune disease scleroderma. They also found titin in condensed

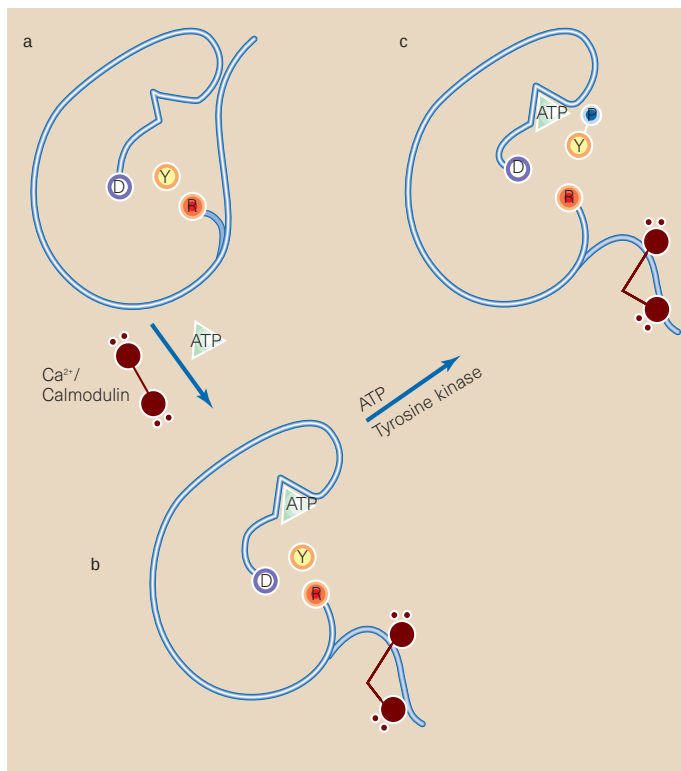


Figure 1 Model for activation of the protein-kinase domain of titin. a, The titin kinase domain is auto-inhibited. Tyrosine (Y170) interacts with the catalytic aspartic acid (D127) and arginine (R129) to prevent binding of the protein substrate. The autoregulatory segment occludes the ATP-binding site. b, Calcium/calmodulin displaces the autoregulatory segment, allowing ATP to bind and exposing Y170. c, An unknown tyrosine kinase phosphorylates Y170, which then interacts with the P+1 residue (R), on the protein substrate.

P+1 loop interacts with the residue immediately carboxy-terminal to the amino acid that will be phosphorylated (P=O) on the protein substrate. Before phosphorylation, the tyrosine-170 residue in the titin kinase domain interacts with the catalytic aspartic acid, as well as with an arginine that is two amino acids carboxy-terminal of this. This interaction blocks the binding site for the protein substrate. Mayans *et al.*¹ now show that phosphorylation of the tyrosine residue relieves this autoinhibition (Fig. 1). Moreover, they surmised that the phosphorylated tyrosine would interact with the P+1 residue of the substrate, and predicted that this residue would be an arginine. Sure enough, when they identified telethonin (an endogenous muscle protein that is a substrate for titin's kinase domain), they found that it contains arginine at P+1. However, this form of autoinhibition is probably rare, because only one other kinase in sequence databases contains the key features — an arginine in the same location relative to the catalytic aspartic acid, and a tyrosine in the P+1 loop.

With the likelihood that titin exists in many, if not all, mammalian cells, and may be involved in both sarcomere formation and chromosome condensation, the new results raise a number of questions. For example, what is the physiological function of the kinase activity, and is it connected to the structural roles of titin? As the telethonin substrate is muscle specific, are there other, biologically relevant, substrates out there? And given that calmodulin binding seems important for kinase activity, even when the enzyme is phosphorylated at a tyrosine residue, does activation of the kinase depend

on a sustained increase in the intracellular concentration of calcium? Will dephosphorylation of tyrosine or a decrease in calcium be enough to inactivate the kinase?

Finally, what about the tyrosine kinase that phosphorylates the titin kinase domain? Titin and calmodulin co-localize to the actin-based cytoskeleton, so is the tyrosine kinase also to be found on the cytoskeleton? A number of the non-receptor tyrosine kinases are localized there¹⁰. Both calcium and tyrosine-kinase activities are frequently increased or deregulated in transformed cells, so is unscheduled activation of titin kinase a cause or a consequence of transformation? If so, could this lead to rearrangement of the actin-based cytoskeleton, which is characteristic of many transformed cells? The visage of titin kinase revealed by Mayans *et al.* will undoubtedly provide new impetus to understand how regulation of this giant protein kinase creates cellular order from potential chaos. □

Anthony R. Means is in the Department of Pharmacology and Cancer Biology, Duke University Medical Center, Box 3813, Durham, North Carolina 27710, USA.
e-mail: means001@mc.duke.edu

1. Mayans, O. *et al.* *Nature* **395**, 863–869 (1998).
2. Labeit, S., Kolmerer, B. & Linke, W. A. *Circ. Res.* **80**, 290–294 (1997).
3. Erickson, H. P. *Science* **276**, 1090–1092 (1997).
4. Machado, C., Sunkel, C. E. & Andrew, D. J. *J. Cell Biol.* **141**, 321–333 (1998).
5. Kemp, B. E. *et al.* in *Protein Kinases* (ed. Woodgett, J.) 30–67 (IRL Press, Oxford, 1994).
6. Kobe, B. *et al.* *EMBO J.* **15**, 5810–5821 (1996).
7. Hariababu, B. *et al.* *EMBO J.* **14**, 3679–3686 (1995).
8. Johnson, L. N., Noble, M. E. M. & Owen, D. J. *Cell* **85**, 149–158 (1996).
9. Goldberg, J., Nairn, A. C. & Kuriyan, J. *Cell* **84**, 875–887 (1996).
10. Parsons, J. T. *Curr. Opin. Cell Biol.* **8**, 146–152 (1996).

Daedalus

Photocoiffure

Most furry animals look neat, smooth and appealing. Each hair on the animal's surface grows so as to reach the envelope defined by all the others. A tree achieves this effect optically: an overshadowed shoot grows faster. So Daedalus reckons that each hair is an optic fibre, carrying ambient light at its tip down to its growing root. If it is 'overshadowed' by the other fibres, it is stimulated to grow. When its tip reaches daylight, growth is turned off. Most animal fur has a visible colour, so the light used must be infrared or deep red. (Thus, despite its obvious camouflage advantages, no animal has green fur; it would absorb the crucial radiation.)

Now human hair must be just as transparent to near infrared. A ruby-red laser treatment can be used to destroy unwanted hair; presumably the massively intense pulse coming down the shaft disorganizes the photosensitive follicle. So why does our hair not grow as neat and smooth as animal fur? Daedalus blames hair-cutting. An uncut hair has a tapered tip, giving a graded match between the refractive index of air and the much higher index of the hair. Radiation enters with little loss — some moth eyes have a micro-spiky surface for just this reason. But a cut hair terminates in an abrupt planar discontinuity, an optical mismatch which reflects away most of the radiation. The root is left permanently in the dark, and the hair just grows uncontrolled. Hair that is seldom or never cut (such as that of the eyebrows) is usually neat and even.

So in barbering, as in much else, once we start interfering with Mother Nature, we are forced to go on doing it. Most of us need endless hair-cutting to avoid looking shaggy and unkempt. Daedalus's analysis provides a neat way out. His 'Fluorostyle' hair-spray is an aerosol loaded with an infrared fluorescing agent. Sprayed onto the head, it is absorbed into the most prominent, outer hairs. The fluorescer absorbs visible light, and emits infrared inside the hair; this is transmitted down the shaft and inhibits growth. Shorter, overshadowed hairs are shielded from the aerosol, and grow normally.

Soon the user will have a neat and even thatch. Selective, graded applications of Fluorostyle spray will then shape, sculpt and maintain it to any desired envelope. The happy user will be as smart and trim as our primitive ancestors — whose popular image as shambling, shaggy, matted monsters can now be seen as totally undeserved.

David Jones

Obituary

André Weil (1906–98)

One of the century's most influential pure mathematicians

André Weil, number-theorist and geometer, died at his home in Princeton, New Jersey, on 6 August. Born in Paris in 1906, he manifested precocious mathematical and philological interests. His gifts and those of his younger sister, the philosopher Simone Weil, were fostered by a watchful, vigorous mother, so that, after several years of the best Parisian lycées and excellent tutors, he entered the *École normale supérieure* at the early age of sixteen. Small of stature, light of build and myopic, he remained until very late in life somewhat of an *enfant terrible*, competitive, a little vain, overbearing with colleagues, but with great resources of charm and culture and an intellectual eagerness that never flagged.

While at the ENS he was admitted to the seminar of Jacques Hadamard, whose style — a wide-ranging, penetrating curiosity about all branches of mathematics — Weil attempted to make his own, very largely succeeding. But his thesis, published when he was 22, was suggested, by his own account, not by the seminar but by his study of the masters of the past, especially of Fermat and Riemann. On number theory, the study of solutions of algebraic equations either in whole numbers or in fractions, the thesis contained among other things a theorem, the Mordell–Weil theorem, that remains to this day of central importance.

Meanwhile, convinced that, with the exception of Hadamard and Élie Cartan, his professors in Paris were out of touch with recent mathematics, he travelled, first to Italy and then to Germany. There, at a time when the mathematical world was much more diverse, both scientifically and linguistically, than today, he could become familiar with the algebraic geometry of the Italians and the number theory of the Germans. His greatest achievement was to be the infusion of number theory with modern geometric notions and the concomitant reworking, by him and by those he influenced, of the foundations of algebraic geometry.

When investigating solutions in whole numbers of an algebraic equation, or of several simultaneous equations, the mathematician's first step is often to replace the equations by congruences. This means to search for whole numbers that render the relevant expressions divisible by



some given number, usually prime.

The number of solutions of any given collection of congruences can be counted. It had been suggested, and proved in some cases, that, for congruences defined by one equation in two unknowns and a given prime, this number (together with the number of solutions of related congruences) could be used to form a function with properties similar to those believed to be possessed by the zeta-function of Riemann, a function of a complex variable. The Riemann hypothesis, which concerns the points at which this function is zero, has been for more than a century the most important unsolved problem of mathematics. But there is an easier Riemann hypothesis for congruence zeta-functions, which yields universal estimates of the number of solutions to the congruences.

Weil succeeded in establishing that hypothesis in unusual personal circumstances. A tourist in Finland at the outbreak of the Second World War, Weil chose to remain there and not to return home to fulfil his military obligations. After the Russian invasion of Finland, he was arrested as a suspected spy. Thanks to a prominent Finnish mathematician, he was not summarily executed but deported to Sweden, where he was taken into custody and returned to France. There, he spent several months in a Rouen prison before being convicted of failure to report for duty. He agreed to serve in a combat unit and the sentence was suspended.

In prison he drafted the main lines of his proof and, circumstances being what they were, quickly published them, but several years elapsed before he completed the prerequisite treatise on the foundations of algebraic geometry, and the reworking, in the context of congruences, of the classical theory of Abelian integrals

and the attendant topological notions on which the proof was based.

These investigations were completed in the United States and in Brazil, where, his combat unit having earlier been evacuated to England after the Franco-German armistice, he spent the war years with his wife and family. Not long after the war, in Chicago, idly reading the papers of Gauss, Weil was led to the general conjectures on congruences and zeta-functions that bear his name. These conjectures have since been demonstrated by Alexandre Grothendieck and his school, and their influence, on Weil himself, on Grothendieck and on others, has permeated number theory and algebraic geometry.

The initial impulse to the recent proof of Fermat's theorem was the insight that it was a consequence of a conjecture of Taniyama, itself prompted by a conjecture of Hasse and Weil. The ideas and techniques that were then used in the proof of Fermat's theorem were many and varied and arose from a multitude of often distant sources, but Andrew Wiles's achievement is unimaginable without notions central to the proof of the Weil conjectures.

There is no question of doing justice here to Weil's manifold contributions to mathematics, but his influence was by no means a result of his theorems alone. A rich and solid, although occasionally strained, prose style in both French and English won a wide audience for his strong views about the nature and the teaching of mathematics.

After more than two years in India, to which an early infatuation with Sanskrit had perhaps drawn him, Weil had returned in 1932 to France, where he and several friends, dissatisfied with the calculus texts available, undertook a collective project that gave birth to an unusual mathematical personality — Nicolas Bourbaki, articulate advocate of a specifically French view of mathematics, author of a widely read treatise in many volumes on its elements, and founder of a seminar that continues to this day.

In 1958 Weil joined the Faculty of the Institute for Advanced Study in Princeton, where he was active until a very few years before his death, as a mathematician and later as an historian of mathematics. His memoirs, written after the death of his wife, to whom he had been deeply attached, throw considerable light on his character.

Robert P. Langlands
Robert P. Langlands is at the Institute for Advanced Study, Olden Lane, Princeton, New Jersey 08540, USA.
e-mail: rpl@ias.edu

Laudable labs?

You can read much about the history of science and of architecture in the changing styles and materials used in the building of laboratories. It's a story of fashion, functionality and financial constraints.

Martin Kemp

We are all too familiar with the messy clutter of disparate laboratory buildings squeezed into congested university campuses. The lab is a major building type, yet we have come to expect little of it — other than as providing functional spaces which almost invariably prove to be inadequate as soon as they are occupied. It would be better if we cared more about the buildings' effects on our visual ambience.

All the larger structures have assumed their external appearance and internal configurations as the result of intense efforts by intelligent patrons and architects, and the older sites embody in microcosm the histories of nineteenth- and twentieth-century architecture. Two telling examples, cheek-by-jowl on the Downing site at Cambridge University, are the School of Agriculture of 1909–10 (now housing the genetics department) and the Wellcome Wing built for the biochemistry department as an extension to the Sir William Dunn Building in 1961–63.

The agriculture school, designed by Arnold Mitchell, proudly proclaims its status through Edwardian rhetoric, above all in the two protruding bays topped by curved pediments, one of which once provided the main entrance. The sculptural adornment includes a proud array of heraldic shields (including the royal arms) and interlaced initials, whose significance has long since faded. In carved swags on either side of the upper windows, a series of agricultural tools — all romantic hand tools and no machinery — are suspend-



The Downing site at Cambridge: Arnold Mitchell's School of Agriculture (now the genetics department), 1909–10 (right), and the Wellcome Wing, 1961–63, designed by Roger Norton for Hammet and Norton.

ed from a sheaf of corn and a ram's head. The motto above the door reads, "Unto God Only Be the Honour and the Glory". The effort and expense devoted to such iconographical display was unthinkable only 50 years later.

The biochemistry wing exudes 1960s functionalism of mass production. Envisaged in principle in 1920, it was not until 1955 that the plans moved towards reality, when the Wellcome Trust offered £120,000 (US\$205,000). This was supplemented by £200,000 from the University Grants Committee in 1957, when Roger Norton was appointed as architect. Beset with the plan-

ning difficulties, delays and arguments that are endemic in committee-run projects, building work finally commenced in 1961.

The Wellcome Wing was a pioneering structure for the site. Built with precast concrete frames and hammered concrete columns, and clad in precast concrete slabs faced with granite aggregate, it belonged very much to the brave new world. Its rhetoric spoke of function, structural integrity, delight in concrete and glass, and modular production — nicely in tune with biochemistry as the study of the modular engineering of organisms. The prominent bicycle shed speaks of utility, while the entrance modestly punctuates the end bay with no more fuss than the windows above.

The building was well received. Nikolaus Pevsner, writing in his canonical *The Buildings of England* in 1970, hailed it as "clear, precise and well detailed ... probably the best building on the whole site". Pevsner, the author of *Pioneers of Modern Design*, was committed to the promotion of modernism.

Almost 40 years after its construction, distance has not lent enchantment to Norton's building. The School of Agriculture has mellowed nicely. The Wellcome Wing has simply aged. Its precast materials have weathered gracelessly. Or is it simply that inadequate distance lends unjustified disenchantment? □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.



Making a statement: the School of Agriculture.



Modest and utilitarian: the Wellcome Wing.

Maternal antibodies block malaria

Women are at increased risk from malaria during pregnancy, and, for unknown reasons, this risk is greatest during the first pregnancy¹. *Plasmodium falciparum*, the most virulent of the four malaria parasites of humans, adheres to a molecule called chondroitin sulphate A (CSA) on the surface of syncytiotrophoblasts (cells lining the intervillous space)² and sequesters in the human placenta. Here we show that anti-adhesion antibodies, which limit the accumulation of parasites in the placenta, appear in pregnant women from Africa and Asia who have been pregnant on previous occasions (multigravidas), but not in those who are pregnant for the first time (primigravidas). Anti-adhesion antibodies against CSA-binding parasites are strain-independent and are associated with greatly reduced prevalence and density of infection. We conclude that malaria susceptibility in primigravidas is related to the lack of these anti-adhesion antibodies, and that an anti-adhesion vaccine for maternal malaria may be globally effective.

We incubated Kenyan parasites in sera from a holoendemic area of western Kenya, then measured the level of parasite binding to CSA. Sera from primigravid women and males failed to inhibit parasite binding (compared with sera from naive controls), whereas sera from secundigravid and multigravid women significantly inhibited binding (Fig. 1a; $P < 0.0001$, compared with primigravid sera). None of the sera inhibited parasite adhesion to CD36, which is the most common receptor for isolates obtained from non-pregnant donors (sera from 18 primigravidas, 17 secundigravidas and 17 multigravidas were tested; data not shown). Purified IgG from five multigravid sera inhibited adhesion to CSA, whereas that from three primigravid samples did not.

We concluded that anti-adhesion activity against CSA-binding parasites is an acquired humoral response, and examined associations with malaria infection. Primigravidas (121 of 382 pregnant women infected; mean, 7.3% parasitized red blood cells in the placenta), who lack anti-adhesion antibodies, were more likely to have placental infections (contingency table analysis, $P = 0.0001$ – 0.0006) and significantly higher levels of parasitaemia (Kruskal–Wallis, $P < 0.0001$) than secundigravidas (41 of 209 infected; mean parasitaemia, 1.5%) and multigravidas (46 of 325 infected; mean parasitaemia, 1.9%). Further, anti-adhesion activity was significantly higher in uninfected than in infected secundigravidas from Kenya (Fig. 1b). Only 14.3% of secundigravidas with serum

anti-adhesion activity that reduced binding to below 35% of control ($n = 49$) were infected at delivery, whereas 57.8% without anti-adhesion activity ($n = 19$) were infected ($P = 0.006$, contingency table analysis). The Kenyan data support the idea that anti-adhesion antibodies protect women from malaria infection during pregnancy.

We then examined parasites and sera from other countries. Three parasite isolates from the peripheral circulation of pregnant Karen women on the Thai–Burmese border bound to both CSA (mean number of parasites bound per 20 high-power fields, 172) and CD36 (mean bound, 155) compared with control BSA (mean bound, 6).

We compared the ability of sera to inhibit Kenyan or Thai–Burmese parasites. In assays measuring parasite binding to CSA, anti-adhesion activity in Kenyan sera was strain-independent; sera from multigravidas inhibited isolates from both countries, whereas sera from primigravidas failed to inhibit isolates from either country (Fig. 1a, c). In serum from pregnant women in the Shire Valley, a holoendemic area of Malawi, anti-adhesion activity was again strain-independent (Fig. 1a, c).

Anti-adhesion activity in sera from Karen donors was absent among primigravidas and variably present in secundigravidas and multigravidas (Fig. 1a, c). The acquisition of anti-adhesion activity at the Thai–Burmese border was delayed compared with African areas of high transmission, probably because the Karen women were infected less frequently than the African women³. For each country studied, sera inhibited the binding of Kenyan and Thai–Burmese isolates to a similar degree (Spearman's rank correlation using results of all tested sera, $\rho = 0.637$, $P < 0.0001$). The CSA-binding parasite is therefore a common feature of maternal malaria globally, and the putative ligand (PfCSA-L) expressed by the parasite to bind CSA is antigenically conserved.

We examined antibody-mediated agglutination of parasite-infected red blood cells. Most of the Kenyan parasite isolates did not agglutinate in the presence of any of the Kenyan sera. When stratified by the presence or absence of agglutinating activity (Fig. 1d), anti-adhesion activity did not vary significantly, suggesting that agglutination and inhibition of adhesion to CSA are

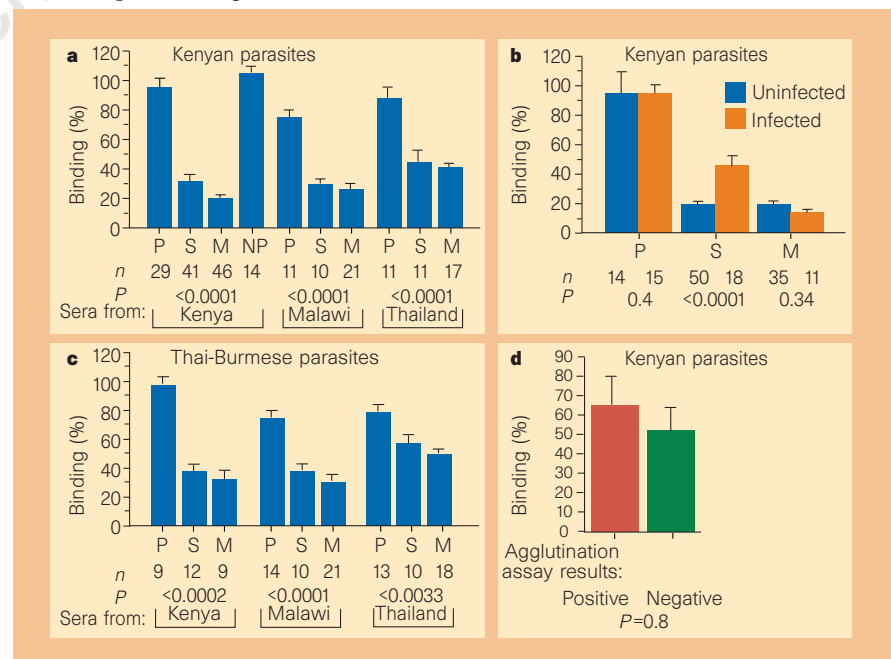


Figure 1 Serum inhibition of malarial parasite binding to CSA, expressed as the percentage of binding in control serum. **a, b, d**, Kenyan parasites; **c**, Thai–Burmese parasites. Parasite isolates collected from placentas in Kenya (**a**) or the peripheral blood of pregnant women in Thailand (**c**) were preincubated with sera from Kenya, Malawi and Thailand (Thailand refers to samples collected from Karen women at the Thai–Burmese border). Kruskal–Wallis analysis was used. **b**, Inhibition of binding using Kenyan sera and parasites, segregated by placental infection status. Mann–Whitney analysis was used. **d**, Inhibition of Kenyan parasite binding with sera donated by Kenyans (5 primigravidas, 3 secundigravidas, 12 multigravidas and 5 males), segregated by results of simultaneous agglutination assay. Paired sign analysis was used (results obtained with the same serum sample against different parasites were paired). P, primigravidas; S, secundigravidas; M, multigravidas; NP, males. Standard error bars are shown. Age of Kenyan donors, as mean (range): primigravidas, 20 (18–26); secundigravidas, 21.5 (18–34); multigravidas, 26.6 (18–36).

distinct phenomena. Antimalarial antibodies defined by agglutination, enzyme-linked immunosorbent assay (data not shown) or surface immunofluorescence (data not shown) were not associated with the presence or absence of infection.

Anti-adhesion antibodies against CSA-binding parasites differ from antibodies against parasites with other binding phenotypes. In studies of CD36-binding parasites, sera were variably effective⁴⁻⁶ and blocked adhesion strain-specifically; the inhibition of adhesion may have been associated with agglutination, which can interfere with adhesion. Our data indicate that anti-adhesion antibodies against CSA-binding parasites uniformly develop in multigravid women from holoendemic areas, appear in Asian and African women, block adhesion in a strain-independent manner, and are distinct from agglutinating antibodies.

We have shown that women with anti-adhesion antibodies against CSA-binding parasites are protected from malaria infection during pregnancy. Because women have limited exposure to the CSA-binding parasite before first gestation, primigravidae lack anti-adhesion antibodies against this parasite subpopulation. We propose that this absence of anti-adhesion antibodies increases susceptibility to malaria during first pregnancies. Further research is required to examine the association between anti-adhesion antibodies and pregnancy outcome, and, most important, to identify PfCSA-L as the next step towards developing a vaccine. An anti-adhesion vaccine that conferred protection against maternal malaria would benefit millions of pregnant women and infants in the tropics, and should be a public health priority.

Michal Fried*, François Nosten†, Alan Brockman‡, Bernard J. Brabin‡, Patrick E. Duffy*

*Department of Immunology,
Walter Reed Army Institute of Research,
14th and Dahlia Street,
Washington DC 20307, USA

e-mail: friedm@wrsmtm-cmail.army.mil

†Shoklo Malaria Research Unit,

PO Box 46, Mae Sot, 63110, Thailand;

Faculty of Tropical Medicine,

Mahidol University, Bangkok, Thailand;

and Centre for Tropical Medicine,

Nuffield Department of Medicine,

John Radcliffe Hospital, Oxford OX3 9DU, UK

‡Tropical Child Health Group,

Liverpool School of Tropical Medicine,

Pembroke Place, Liverpool L3 5QA, UK

1. McGregor, I. A. *Am. J. Trop. Med. Hyg.* **33**, 517–525 (1984).

2. Fried, M. & Duffy, P. E. *Science* **272**, 1502–1504 (1996).

3. Nosten, F., ter Kuile, F., Maelankirri, L., Decludt, B. & White, N. J. *Trans. R. Soc. Trop. Med. Hyg.* **85**, 424–429 (1991).

4. Singh, B. et al. *Clin. Exp. Immunol.* **72**, 145–150 (1988).

5. Southwell, B. R. et al. *Trans. R. Soc. Trop. Med. Hyg.* **83**, 464–469 (1989).

6. Udeinya, I. J., Miller, L. H., McGregor, I. A. & Jensen, J. B. *Nature* **303**, 429–431 (1983).

Object recognition can drive motion perception

When two spatially separated spots of light are flashed in rapid succession, apparent motion is seen between them¹⁻³. We extended this phenomenon by photographing a face and producing from it a fragmented 'puzzle picture' or 'Mooney face' in which the face is not initially visible (Fig. 1, left; frame 1) but is seen after 15 to 60 seconds⁴. Another photograph of the same face seen in profile was used to produce a second Mooney face (Fig. 1, right; frame 2). When the two images were alternated, naive subjects at first saw random, incoherent, two-dimensional (2D) motion between the fragments. But once the face was recognized, it was perceived to rotate unambiguously in three dimensions. We conclude that complex image tokens set up by perceptual learning can drive perception of apparent motion.

To what extent are different aspects of perception mediated by specialized modules⁵⁻⁸ in the brain? Attention can influence the way in which corresponding points are seen to move³ and, in ambiguous, bi-stable displays, people tend to see 'plausible' trajectories (for example, a bomb is more likely to be seen going down rather than up⁹). But it is usually assumed that these are weak second-order effects, no more than a small bias in the perceived direction of motion.

Cells in area MT of the monkey brain respond to apparent motion¹⁰ and there is psychophysical evidence that primitive, rather than sophisticated, visual features are used for motion correspondence¹¹. As a result, most computational models of motion perception emphasize a stronger modular approach^{12,13}, arguing that motion correspondence is not influenced by high-level object knowledge and semantics. For instance, it is easy to see a pig's face transforming into a human face even though this is logically absurd.

Here we use a new type of apparent-motion display to investigate directly how perceptual learning and high-level object knowledge are involved in motion processing. First, we illuminated the face of a model looking at the camera lens with a single strong light source. We then photographed the face and digitized the image to obtain a fragmented puzzle picture (Fig. 1, left). The model then rotated his face by 45 degrees (or 90 degrees) and a second image was made using the same procedure (Fig. 1, right). A set of four pairs of such images was made of four different faces. The direction of illumination and degree of head rotation were chosen to eliminate local matches between the shadows and to optimize the difference between the appearance of the images before and after perceptual learning⁴.

Naive subjects could often not see the faces in these images. Of 28 subjects, 17 could not see the face initially in at least one of the four sets of photographs, and four of these subjects could not do so in two. These 17 volunteers were first shown various demonstrations of apparent motion (but not of faces) to familiarize them with the illusion. They were then shown frames 1 and 2 of the images in which no face was perceived, alternated at 3 Hz (ISI = 0) for 20 seconds, and asked to describe the perceived direction of motion, if any. The subjects saw either random chaotic motion or motion along arbitrary 2D trajectories that often varied from trial to trial (in 20 of 21 experimental sessions). When asked explicitly if they saw any 3D motion, 16 subjects said they had not; the seventeenth subject immediately saw a face rotate in one of the displays as soon as the two frames were alternated.

The film sequence was then stopped and the same subjects were asked whether they could see a human face, a prompt that always resulted in the face being seen after 10 to 30 seconds (we often had to resort to pointing to individual features after 30 seconds). Remarkably, when the same two frames were alternated, all subjects in all



Figure 1 Two views of a 'Mooney'-type face produced by digitizing a video image. Left, frame 1; right, frame 2. A demonstration can be produced by showing the two frames alternately as an upside-down movie; nearly all subjects see 2D movement. But if the same movie is viewed upright or the subjects are told it is a face, they invariably see rigid 3D rotation.

trials saw 3D rotation of the face, from frontal to semiprofile or profile. No 2D expansion, contraction or chaotic incoherent motion was ever seen. Because the display was exactly the same but the perceived motion trajectory had changed completely, we conclude that high-level object representations can drive apparent motion. In particular, a temporary 'object' (or extended contours forming portions of it) created exclusively by perceptual learning can provide an input for apparent motion. The meaning of the perceived configuration therefore helps to eliminate an infinite set of false matches between the fragments, allowing motion correspondence to be established correctly between the two views of the face.

Similar effects were seen with other puzzle pictures, such as a hidden dalmatian dog¹⁴, with the dog sitting and standing in alternate frames. When the two frames were alternated, subjects initially saw random 2D motion, but once the dog was seen it was perceived to sit and stand alternately. This effect is particularly remarkable because the second image was made using a completely different dog with different spots, so no rigid transformation was possible between the two images.

Taken together, these observations suggest that strong interactions occur between perceptual modules concerned with motion (for example, in area V5)⁵ and those involved in high-level object recognition (for example, in area IT)⁴. The strictly hierarchical, modular models of perception popularized by computer scientists should therefore be replaced by a more dynamic view. From every stage in the hierarchy, partial answers may be sent back to bias the processing in earlier stages.

V. S. Ramachandran, C. Armel, C. Foster, R. Stoddard

Center for Brain and Cognition,
University of California at San Diego,
La Jolla, California 92093, USA
e-mail: vramacha@ucsd.edu

- Helmholtz, H. L. F. von *Handbuch der Physiologischen Optik* (Leipzig, 1867).
- Kolers, P. *Aspects of Motion Perception* (Pergamon, New York, 1972).
- Ramachandran, V. S. & Anstis, S. M. *Sci. Am.* **254**, 102–109 (1986).
- Tovee, M. J., Rolls, E. T. & Ramachandran, V. S. *Neuroreport* **7**, 2757–2760 (1996).
- Zeki, S. M. *Nature* **274**, 423–428 (1978).
- Allman, J. in *Oxford Companion to the Mind* (ed. Gregory, R. L.) (Oxford Univ. Press, 1987).
- Livingstone, M. S. & Hubel, D. H. *J. Neurosci.* **7**, 3416–3468 (1987).
- Van Essen, D. C. *Annu. Rev. Neurosci.* **2**, 227–263 (1979).
- Bruner, J. *Psychol. Rev.* **64**, 123–152 (1957).
- Mikami, A., Newsome, W. T. & Wurtz, R. H. *J. Neurophysiol.* **55**, 1308–1327 (1986).
- Ramachandran, V. S. & Inada, V. *Spatial Vis.* **1**, 57–67 (1985).
- Marr, D. *Vision* (Freeman, San Francisco, 1982).
- Ullman, S. *The Interpretation of Visual Motion* (MIT Press, Cambridge, MA, 1979).
- Gregory, R. L. *The Intelligent Eye* (McGraw Hill, New York, 1970).

Preserving tardigrades under pressure

When an animal is exposed to high hydrostatic pressure, its cellular membranes^{1,2}, proteins^{3–5} and DNA⁶ are damaged. At pressures of around 30 megapascals (MPa), proliferation and metabolism in microorganisms stops; at 300 MPa, most bacteria and multicellular organisms die. But here we show that, in perfluorocarbon at pressures as high as 600 MPa, small terrestrial animals known as tardigrades can survive in a dehydrated state.

Terrestrial tardigrades become immobile and shrink into a form known as the 'tun' state when the humidity decreases. In this state, they can survive extreme temperatures, as low as –253 °C or as high as 151 °C, as well as exposure to a vacuum or to X-rays^{7–10}. We have now tested the ability of the tardigrades *Macrobiotus occidentalis* (order Eutardigrada) and *Echiniscus japonicus* (Heterotardigrada) to survive under extraordinarily high hydrostatic pressures.

We sealed *M. occidentalis* tardigrades in a small plastic container (6 ml) placed inside a pressure capsule (R7K-3-10, Yamamoto Suiatu Kougyousho) and compressed using water as the pressure medium. The outside temperature was 21 °C and the water temperature inside the capsule was 25 °C; stepped hydrostatic pressures were applied for 20 minutes at a time (100, 200, 300, 400, 500 and 600 MPa). Pressure was increased by 100 MPa per minute and then decreased at the same rate. After decompression, *M. occidentalis* was removed from the capsule

and examined under a light microscope, which revealed that all organisms died at pressures over 200 MPa (Fig. 1).

We then investigated whether tardigrades could acquire pressure resistance in the tun state (a process known as anhydrobiosis) by dehydrating them before applying pressure. *M. occidentalis* and *E. japonicus* were dehydrated on filter paper in Petri dishes for more than 24 hours, when the relative humidity in the dishes dropped from 70–80% to 10–30%. To prevent the tardigrades from rehydrating during compression, we used an inert solvent, perfluorocarbon (C₈F₁₈ Fluorinate PC77, Sumitomo 3M), as the pressure medium instead of water. Tardigrades were then removed from the pressure capsule and soaked in water to rinse off the perfluorocarbon. One hour later, we confirmed that they had changed from the tun state to the active state.

To test whether the perfluorocarbon increased the survival of tardigrades exposed to high hydrostatic pressure, we subjected tardigrades in perfluorocarbon, which were still in the active state, to the same hydrostatic pressure changes. All active-state tardigrades were dead at pressures above 200 MPa.

We evaluated the data at 600 MPa for the group ($n = 20$) in the tun state in perfluorocarbon and in the active state in water and perfluorocarbon. The survival rate of *M. occidentalis* was 95% at 600 MPa (Fig. 1), and there was a difference between active-state and tun-state animals ($P < 0.01$; chi-squared). The survival rate of *E. japonicus* was 80%, as some animals had died and their fluid had leaked onto the filter paper, which we attributed to inadequate

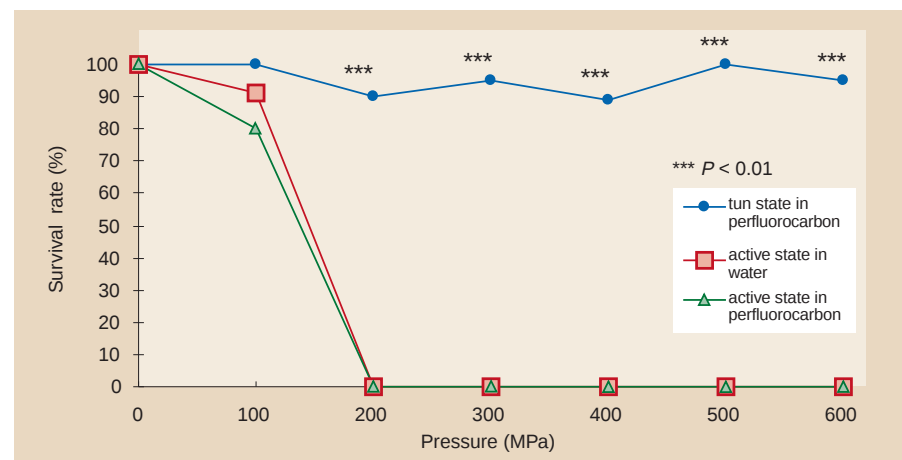


Figure 1 Survival rate of *Macrobiotus occidentalis* after exposure to high hydrostatic pressure. Before compression (100 MPa min⁻¹), tardigrades were either in the tun state (dehydrated) in perfluorocarbon, or in the active state in water or perfluorocarbon. Either water or perfluorocarbon was used as pressure medium. An hour after high hydrostatic pressure was maintained for 20 minutes and decompression to –100 MPa min⁻¹, the animals were soaked in distilled water for an hour, and those in the active state were examined with an optical microscope (magnification, × 40). When animals in the tun state were exposed to 100–600 MPa in perfluorocarbon, a high survival rate was obtained. The survival rate of active-state animals in perfluorocarbon is 80% at 100 MPa and 0% at pressures greater than 200 MPa. This survival rate at 100 MPa was lower than the 91% obtained using water at 100 MPa, so perfluorocarbon did not increase the survival rate.

dehydration before the experiment.

Tardigrades are composed of about 40,000 cells, which survive not only high-speed compression under a hydrostatic pressure of 600 MPa (equivalent to six times the pressure of sea water at a depth of 10,000 metres), but also being maintained at this pressure, and high-speed decompression as well. With perfluorocarbon as the pressure medium, we have demonstrated the viability of tardigrades after keeping them in an anhydrobiotic state. This viability is influenced not just by pressure but by the absolute amount of water in the organism's body, enabling us to exploit its dehydration/water-absorption mechanism for preservation purposes.

Kunihiro Seki, Masato Toyoshima

Department of Biological Science,
Faculty of Science, Kanagawa University,
2946 Tsuchiya, Hiratsukashi 259-1293, Japan
e-mail:sekikuni@heather.ed.info.kanagawa-u.ac.jp

1. Macdonald, A.G. in *Current Perspectives in High Pressure Biology* (eds Jonnasch, H. W., Marquis, R. E. & Zimmerman, A. M.) 207–223 (Academic, London, 1987).
2. Ookuma, M. in *Pressure Processed Food-Research and Development* (ed. Hayashi, R.) 157–164 (San-Ei, Kyoto, 1990).
3. Suzuki, K. *Symp. Soc. Exp. Biol.* **26**, 103–124 (1972).
4. Weber, G. & Drickamer, H. G. *Q. Rev. Biophys.* **16**, 89–112 (1983).
5. Heremans, K. in *Current Perspectives in High Pressure Biology* (eds Jonnasch, H. W., Marquis, R. E. & Zimmerman, A. M.) 224–244 (Academic, London, 1987).
6. Murakami, T. H. & Zimmerman, A. M. *Cytobiosis* **7**, 171–181 (1973).
7. Rahm, P. G. *Ann. Zool. Jpn* **16**, 345–352 (1937).
8. Crowe, J. H. & Cooper, A. F. *Sci. Am.* **225**, 30–36 (1971).
9. May, R., Maria, M. & Gulmard, J. *Bull. Biol. Fr. Bel.* **48**, 349–367 (1964).
10. Kinchin, I. M. *The Biology of Tardigrades* (Portland, London, 1994).

Specific interference by ingested dsRNA

A genetic interference phenomenon in the nematode *Caenorhabditis elegans* has been described in which expression of an individual gene can be specifically reduced by microinjecting a corresponding fragment of double-stranded (ds) RNA¹. One striking feature of this process is a spreading effect: interference in a broad region of the animal is observed following the injection of dsRNA into the extracellular body cavity. Here we show that *C. elegans* can respond in a gene-specific manner to dsRNA encountered in the environment. *C. elegans* normally feed on bacteria, ingesting and grinding them in the pharynx and subsequently absorbing bacterial contents in the gut. We find that *Escherichia coli* bacteria expressing dsRNAs can confer specific interference effects on the nematode larvae that feed on them.

Three *C. elegans* genes were used for this analysis. For each gene, a corresponding dsRNA was expressed in *E. coli* by inserting

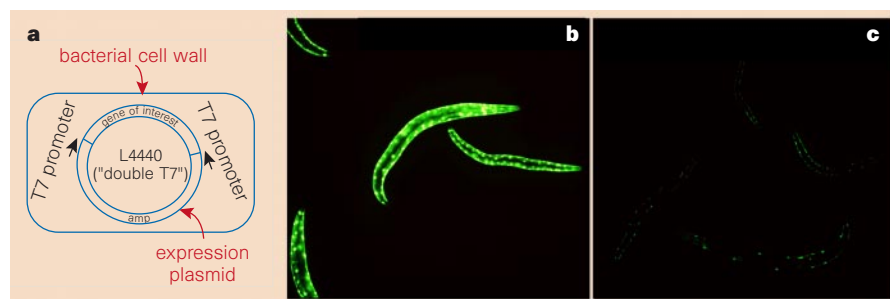


Figure 1 Genetic interference following ingestion of dsRNA-expressing bacteria by *Caenorhabditis elegans*.

a, General scheme for dsRNA production. Segments were cloned between flanking copies of the bacteriophage T7 promoter into a bacterial plasmid vector (pPD129.36; J. Fleenor and A. F., unpublished). A bacterial strain (BL21/DE3; ref. 7) expressing the T7 polymerase gene from an inducible (Lac) promoter was used as a host. As an alternative strategy, we used a single copy of the T7 promoter to drive expression of an inverted duplication for a segment of the target gene (*unc-22* or *gfp*). A nuclease-resistant dsRNA was detected in lysates of these bacteria. The two bacterial expression systems gave similar interference results. **b**, A GFP-expressing *C. elegans* strain (PD4251)¹ fed on a naive bacterial host. Animals show high GFP fluorescence in body muscles. **c**, PD4251 animals reared on bacteria expressing dsRNA corresponding to the *gfp* coding region. Under the conditions of this experiment, 12% of these animals show a dramatic decrease in GFP.

a segment of the coding region into a plasmid vector designed for bidirectional transcription by bacteriophage T7 RNA polymerase. The dsRNA segments used for these experiments were the same as those used previously¹. We then observed the results of feeding these bacteria to *C. elegans*, and compared the effects with those of loss-of-function mutations and to animals microinjected with dsRNA.

The *C. elegans* gene *unc-22* encodes an abundant muscle filament protein². Null mutations³ or injection of *unc-22* dsRNA¹ produce a characteristic and uniform twitching phenotype in which the animals can sustain only transient muscle contraction. When wild-type animals were fed bacteria expressing a dsRNA segment from *unc-22*, 85% exhibited a weak but distinct twitching phenotype characteristic of partial loss of function for the *unc-22* gene.

The gene *fem-1* encodes a late component of the *C. elegans* sex-determination pathway^{4,5}. Null mutations⁴ or injection of dsRNA¹ prevent the production of sperm and lead euploid (XX) animals to develop as females (wild-type XX animals develop as hermaphrodites). When wild-type animals were fed bacteria expressing dsRNA corresponding to *fem-1*, 43% exhibited a spermless (female) phenotype and were sterile.

We then assessed the ability of dsRNA to interfere with a transgene target. When animals expressing a green fluorescent protein (GFP) transgene were fed bacteria expressing dsRNA corresponding to the *gfp* reporter^{1,6}, a decrease in GFP fluorescence was observed in about 12% of the population (Fig. 1).

The effects of bacteria carrying different dsRNAs were fully gene specific: dsRNAs from *fem-1* and *gfp* produced no twitching; dsRNAs from *unc-22* and *gfp* did not produce females; and dsRNAs from *unc-22* and *fem-1* did not reduce GFP expression. These

interference effects were evidently mediated by dsRNA, as bacteria expressing only the sense or antisense strand (for *gfp* or *unc-22*) caused no evident phenotypic effects (data not shown).

As with injected dsRNAs, the effects of feeding dsRNA to *C. elegans* are reversible and do not reflect a stable genetic change, as transfer of affected animals back to non-engineered bacterial food led within a generation to loss of the Unc-22 or faint-GFP phenotypes. From an engineering perspective, bacterial-mediated delivery of dsRNA is less effective than direct microinjection. This is evident from the frequency and severity of the interference phenotypes discussed above, and from observations that, for several genes known to be inhibited by injected dsRNA, bacterially mediated interference was marginal or not evident. Differences in susceptibility could reflect resistance of some cells or stages to the consequences of ingested dsRNA.

This work provides an example of RNA-mediated transfer of information between organisms and between species. It is not yet known whether such RNA-mediated interference-transfer mechanisms participate in natural ecological interactions, such as antiviral defence or communication during biological symbiosis.

Lisa Timmons, Andrew Fire

Department of Embryology,
Carnegie Institution of Washington,
115 West University Parkway,
Baltimore, Maryland 21210, USA
e-mail:fire@mail1.ciwemb.edu

1. Fire, A. et al. *Nature* **391**, 806–811 (1998).
2. Benian, G., L'Hernault, S. & Morris, M. *Genetics* **134**, 1097–1104 (1993).
3. Brenner, S. *Genetics* **77**, 71–94 (1974).
4. Doniach, T. & Hodgkin, J. A. *Dev. Biol.* **106**, 223–235 (1984).
5. Spence, A., Coulson, A. & Hodgkin, J. *Cell* **60**, 981–990 (1990).
6. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. & Prasher, D. *Science* **263**, 802–805 (1994).
7. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Methods Enzymol.* **185**, 60–89 (1990).

Beefing up the food police

Poison on a Plate: The Dangers in the Food We Eat and How To Avoid Them

by Richard Lacey

Metro Books: 1998. 282 pp. £17.99 (hbk), £12.99 (pbk)

Eat Your Genes: How Genetically Modified Food Is Entering Our Diet

by Stephen Nottingham

Zed Books: 1998. 212 pp. £32.95, \$45 (hbk), £11.95, \$17.95 (pbk)

John Godfrey

For much of *Poison on a Plate*, Richard Lacey describes his part in pointing out the possible danger to man from bovine spongiform encephalopathy (BSE), and how he was ridiculed and persecuted for his pains. He was a better guide than the official Southwood Report of 1989, which stated that "it is likely that cattle will prove to be a 'dead-end host' for the disease agent and most unlikely that BSE will have any implications for human health". He has partly suffered, though, for his sporadic lack of wisdom. For example, he gives no basis for his "grim estimate of a possible five per cent death rate amongst the population of the United Kingdom within the next ten years or so".

Lacey may well have been in the mind of Sir Robert May, the Government's chief scientific adviser, when he recommended recently that his colleagues "consult widely, and consult contrary opinion" before issuing advice (*Nature* 393, 720; 1998). The effectiveness of the Food Standards Agency now being proposed for the United Kingdom depends on the full and prompt availability



Cow to the slaughter: the BSE crisis prompted moves towards creation of a UK Food Standards Agency.

of critical information. In complex problems like those of BSE nobody can judge who needs to know what crucial fact, so the only safe course is for information to be freely available to anyone who wants to know it. The parliamentary bills to set up the food agency and to provide freedom of information will work best together.

Lacey is also justified in pointing out that he was proposing a food watchdog long before the UK government accepted the idea. As it is now conceived, the Food Standards Agency will be devoted to the needs of people as consumers of safe food of high nutritional quality, and it will be independent of producers' interests. The proposal has been widely welcomed. It will be in the Queen's speech for the coming parliamentary session, as long as the government considers it sufficiently urgent. So BSE has had some positive consequences.

Neither of these books mentions that the European Union (EU) has reacted in a similarly helpful way. The BSE crisis has brought changes to the balance of power between the union's institutions. The European Parliament was critical of the United Kingdom, but perhaps more importantly it castigated the European Commission severely for not implementing adequate control measures promptly. Last year, for the first time in its history, the parliament threatened to use its power to sack the commission, to force it to make unprecedented changes. The Directorate General for Consumer Affairs has been greatly strengthened, and its advisory committees have been made more independent. This directorate is now using the pow-

ers it already had, but previously failed to use, to get involved in the policies of other directorates general, particularly that for agriculture. Much of the United Kingdom's domestic food safety, including the safety of traded meat, is now regulated at an EU level. So the reforms are a healthy development for all member states.

The European Parliament is set, from now on, to be a more effective watchdog not only on BSE, but in the public interest generally. The new Scientific Steering Committee of the European Commission has set up a study of three categories of risk of BSE for specified geographical areas throughout the EU. Some member states still allow mammalian meat and bone meal to be fed to animals other than ruminants, which means that this risky feed is present on farms, and may well be eaten inadvertently by cattle and sheep.

The last three chapters of *Poison on a Plate* are less personal. They mainly provide practical advice that is decidedly helpful to anyone wanting to reduce the risk of getting food poisoning. However, on a nutritional point I am mystified that he castigates crisps because of their fat, but makes no criticism of fish and chips on the same grounds. He is unfair to the UK National Consumer Council. It is not under the wing of the agriculture ministry as he says, but gets its government funds from the Department of Trade and Industry. He just touches on genetic engineering, but loses his biological judgement. He illustrates the technology by writing that it will soon be in the power of man to genetically engineer a cow with flesh that tastes like chicken or even fish. So far reasonable



Polarized debate: the issue of genetically modified food is raising strong passions.

NICK COBBING/STILL PICTURES

NIGEL DICKINSON/STILL PICTURES

enough, but he goes on to state that such cows could grow to the size of an elephant. He seems to have forgotten the inverse square rule. Such a cow would collapse with snapped legs, lacking the elephant's profound evolutionary reorganization to cope with its weight.

Stephen Nottingham's *Eat Your Genes* gives a very useful account of the regulatory process for genetically modified (GM) foods in the EU and the United Kingdom, so it is odd that he seems unaware of how much change has recently taken place in Europe. Perhaps more seriously, he barely mentions the *Codex Alimentarius* Commission, set up by the United Nations in 1962. Consumers, as opposed to producers, have very little say in its operation; it too needs a decisive shift in the balance of power within it towards the public interest.

The EU and the United States are at odds over the safety of traded food in general but of GM foods in particular, as are the member states of the World Trade Organization. Nottingham is interesting on the increasing political influence of multinational companies on the WTO. Reforms in the interest of consumer safety, similar to those in train for the United Kingdom and the EU, are urgently needed at the global level.

The public has rapidly become more

interested in the use of genetic modification in food production, and Nottingham's book deserves to be widely read. Most of the debate is now highly polarized, either for or against the technology in principle. *Eat Your Genes* is refreshing, if less entertaining as a spectator sport, for being descriptive rather than particularly partisan. He is good on the possible dangers from gene combinations that would not have been produced by traditional selective breeding. Ordinary hazard analysis depends on a track record of past accidents to calculate future risks. Modern biotechnology is just too new for such calculations to be realistic.

Nottingham went to press too soon to welcome the Royal Society's statement "Genetically modified plants for food use". It recommends that the government commission an over-arching body to bring together a number of departmental responsibilities and to monitor the wider issues associated with GM plants. The statement lists problems that this new body should consider if the existing advisory committees do not, such as long term impacts on the ecosystem.

Nottingham is right to emphasize that most developments so far have been driven by profit rather than consumer needs. He could usefully have written more on long-term problems that need public funding for a

search for solutions. Take the food supply if the climate changes. If the climate becomes more extreme, agriculture will need varieties that can cope. As global warming becomes an acute problem, the market mechanism will work; but it will by then be too late for anything close to an optimal response to the problem. Even if the climate does not change much, research that adapts crops and farm animals to harsh conditions would be valuable in those parts of the world already short of food, and which will be shorter still with the increase in world population. Another problem is the nutritional quality of foods such as vegetables or fruits. The relative quantities of vitamins or micronutrients that they contain are not apparent to consumers as they buy a particular variety, so here the market mechanism is less effective. It is of course more efficient for qualities such as appearance or shelf life.

In discussions that ought to centre on a cost-benefit analysis of GM foods, the critics focus on the costs, the industry on the benefits. Nottingham's book will help his readers to make the debate more fertile — and less boring. □

John Godfrey is at 41 Lawford Road, London NW5 2LG, UK.

Hawaiian headgear

These feather 'god heads' were brought to England from the Hawaiian islands on one of Captain James Cook's ships in 1780, and painted by Sarah Stone. The artist was employed by the entrepreneur Sir Ashton Lever to record the 'curiosities' in his private museum. Besides ethnographic material collected on Cook's round-the-world voyages, the museum contained specimens used by naturalists for their descriptions of new species. From *Sarah Stone: Natural Curiosities from the New Worlds* by Christine E. Jackson (Merrell Holberton/Natural History Museum, London, £29.95, \$45). The book is in the new Art of Nature series which highlights little-known artists whose work is preserved in the Natural History Museum.



The freedom not to listen

Silencing Scientists and Scholars in Other Fields: Power, Paradigm Controls, Peer Review and Scholarly Communication

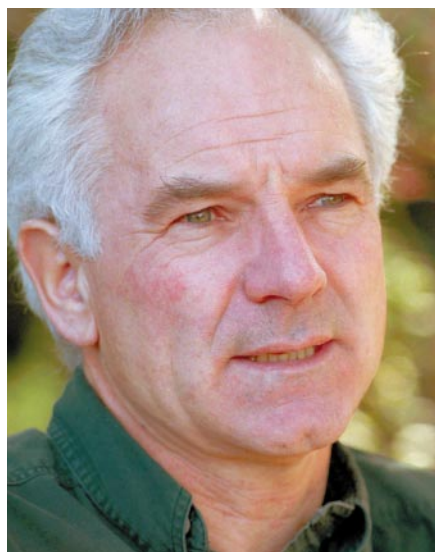
by Gordon Moran

JAI Press: 1998. 186 pp. £47, \$73.95 (hbk), £25, \$39.50 (pbk)

John Ziman

We come into science starry-eyed, upright and trustful of our fellow pilgrims in this noble venture. Alas, we and they prove as morally frail as the rest of humanity. In time, we mostly learn to accept and cope with the petty injustices, vanities and deceptions incidental to our profession. But occasionally an apparent affront or iniquity cannot be tolerated, and bursts out of the private sphere into public conflict. And very occasionally a contestant seeks support by generalizing their case into an attack on some feature of the whole system.

Gordon Moran's argument is that he, and many other scientists and scholars, have been "silenced" — that is, they have been deliberately prevented from saying or writing something that ought, in all conscience, to have been said or written on some particular scientific or scholarly matter. In his case, the contested point was whether the great equestrian portrait of Guido Riccio in Siena was actually painted in 1330 by Simone



Peter Duesberg's claims that HIV is not the cause of AIDS have not been censored but simply rejected.

Martini, or by a less famous artist in 1352. I have no idea of the rights or wrongs in this controversy, but Moran's account of it is so obviously one-sided that I discount it as evidence of a more general phenomenon.

Indeed, his definition of this phenomenon is altogether too general. The case of the scientist prevented by the body that employs him from revealing an embarrassing research result is entirely different from that of the academic scientist whose paradigm-breaking discovery, although widely publicized, is ignored by an intellectually blinkered scientific community. At one end of the scale we may find illicit intimidation to cover up a crime; at the other end, a substantial group of free citizens has a perfect right not to be bounced by a zealous heretic out of what they honestly consider to be the established truth, however wrong-headed this may appear in hindsight.

By fudging the notion of "whistle-blowing" to include cases that are really quite different — for example, Peter Duesberg's well known but thoroughly controverted claims about AIDS — Moran also misses another very important point. On the one hand, it is truly scandalous that scientists employed in confidential research can be muzzled, against the public interest, by quasi-legal corporate coercion. On the other hand, contractual confidentiality is a necessary element of commerce, and prudent silence can sometimes be a civic virtue. For example, an unjustified public accusation of scientific fraud is as reprehensible, and as punishable in law, as any other irresponsible defamatory communication. But this ethical ambivalence is not specific to science, since it clearly afflicts many other trusted professionals such as accountants, engineers, physicians and even security guards.

The heart of the matter, which Moran skirts around but does not tackle directly, is

whether a scientist has an intrinsic right to voice unwelcome scientific opinions in public. The legal situation is quite clear. In a democratic society, scientists have exactly the same rights to free speech as other citizens — no more and no less. They are perfectly at liberty to present their views vocally or in written form. Of course, this might be very costly, but so it is for people or organizations with religious, political or commercial claims that they believe to be just as cogent.

In practice, of course, what these "scientists and scholars in other fields" really want is to be permitted to present their views in reputable journals and books, preferably for free. But here authors have limited legal leverage, except in respect to copyright. The huge social institution called 'the scholarly literature' is largely controlled by unpaid editors and referees and is regulated almost entirely by customary practices and commercial opportunities. It is not answerable to any other authority for its decisions, and cannot easily be brought to book for its occasional follies and injustices.

On the whole, scientific and scholarly editors are extremely conscientious, but so much has to depend on opinion. As Moran admits in relation to the Internet, without critical filtering all serious scientific discourse would be drowned in a flood of wild shouting. But, since editors have no alternative to peer review, how can they ensure that their advisers are not biased? Since expert referees often fail to understand unconventional arguments, who might teach them to be more open-minded? And since the scientific archives are riddled with entrenched errors, how could they ever be intellectually cleansed?

In the end, however, I am not persuaded that the cases Moran cites add up to a systematic pattern of injustice. Let us remember that the scientific community is an agonistic forum. Knowledge is created as much by heated argument as by ice-cold experimentation. The norms and practices of scientific communication pit researchers verbally against one another, but strictly limit their rhetorical weapons. Whatever the hidden passions, superficial impersonality, courtesy and respect for past achievement are as mandatory in the public arena as impeccable logic and empirical fact. It may be deeply hurtful to be defeated in such combats, but it is not necessarily unjust or even shameful. One may yearn to have one's ideas accepted and acclaimed, but one has to accept that the world may decide differently.

But this book is a timely reminder to all of us in the academic community that the practices that are the prime guarantee of the credibility of our enterprise can seem very cruel to those who are crushed by them. □

John Ziman, emeritus professor of physics at the University of Bristol, is at 27 Little London Green, Oakley, Aylesbury, Bucks HP18 9QL, UK.

How the brain holds our attention

The Attentive Brain

edited by Raja Parasuraman
MIT Press: 1998. 577 pp. \$65, £51.95

Steven Yantis

In the last decade, explosive growth in the field of cognitive neuroscience has yielded a new scientific society, several new journals and hundreds of articles (many in these pages), all focused on how the brain subserves perception, cognition and behaviour. This convergence of methodologies from psychology, neuroscience, neuroimaging and cognitive neuropsychology has produced dramatic advances in our conception of problems that previously had been addressed separately in each field.

Raja Parasuraman has now compiled an exciting collection of papers aimed at one of the most fruitful subjects cognitive neuroscientists have tackled: attention. The overarching questions are these: given that we can perceive, think about and do only (roughly) one thing at a time — in other words, given that perception, cognition and action are necessarily selective — how is that selection achieved, and how is it coordinated with the behavioural demands faced by the organism? How do behavioural goals, for example, modulate sensory input? How are multiple simultaneous tasks juggled? The consensus that emerges from this book is that there is no single 'centre' for attention in the brain; instead, there are multiple distributed systems of attention that keep things running smoothly and efficiently.

Eight chapters in the first section provide incisive tutorials on the methods of cognitive neuroscience as applied to the study of attention. These methods include invasive neuro-anatomical and neurophysiological techniques, noninvasive neuroimaging approaches (electroencephalography, positron emission tomography and functional magnetic resonance imaging), the analysis of impaired performance in people with brain damage, and computational modelling of attentional control. These chapters alone make the book exceptionally valuable: they offer a comprehensive overview of the key methods of cognitive neuroscience, and their focus on a common topic illustrates how a multifaceted issue like attention can be approached from many different angles, each offering a distinct perspective. This section is aimed at readers who have only a minimal familiarity with basic neuroscience.

Each of the remaining chapters focuses on a substantive question about the workings of attention and then applies one or more of the techniques discussed in the first section to that question. Several of the chap-



Brain waves: "there is no single 'centre' for attention in the brain".

ters in this part of the book employ a method not discussed in the first section, that is, the approach of cognitive psychology. They focus on the analysis of behaviour itself (the speed and accuracy with which perceptual and cognitive tasks are carried out) in populations of normal people, and draw inferences from the behavioural data about how the brain must be organized.

The topics were selected for their value in illustrating cognitive neuroscientific approaches, and the coverage is therefore somewhat idiosyncratic — readers who desire a broader overview of current theoretical issues in attention research may wish to consult offerings in the "Attention and Performance" series (published by Erlbaum and MIT Press), or Harold Pashler's recent tutorial volume, *Attention* (Psychology Press, 1998). However, these chapters provide excellent introductions to more than a dozen fascinating problems concerning how the

brain exerts control over cognition.

One is always tempted to be dazzled by the technological advances that permit increasingly detailed and, in the case of technologies like functional neuroimaging, aesthetically pleasing measurements of brain function. It is these advances, after all, that have made such rapid progress in cognitive neuroscience possible. However, the take-home message of this volume is that satisfactory progress requires careful attention not just to the brain, but to the behaviour that the brain produces. In chapter after chapter, the authors rely on a detailed analysis of the psychological requirements of behavioural tasks to motivate the design and interpretation of their experiments.

Parasuraman's 1984 publication *Varieties of Attention* (with D. R. Davies) was a highly influential collection detailing the problems in attention then faced by cognitive psychologists, and it continues to be widely cited. This new volume promises to be even more useful. □

Steven Yantis is in the Department of Psychology, Johns Hopkins University, Baltimore, Maryland 21218-2686, USA.

Looking after your molecules

The Molecules within Us: Our Body in Health and Disease

by Charles A. Pasternak

Plenum: 1998. 335 pp. £17.55, \$28.95

John Emsley

The traditional groupings in chemistry — organic, inorganic, physical and theoretical — have all but dissolved, due mainly to their interactions with outside disciplines. The demands of those in medicine, biology and pharmacology are changing the foci of chemistry, but this meeting of minds has given rise to the creation and investigation of molecules of fascinating complexity. Their cooperation could also have huge medical benefits.

When science moves quickly and promises to impinge on everyday life, it is important to explain to the public what is going on, before the mischief makers get to work. So far their opposition to the healing sciences has been muted, merely emphasizing the side-effects that new drugs and treatments can have, but of late their voice has become more strident, playing on public ignorance about genetic manipulation and encouraging exaggerated fears about what this might lead to.

For these reasons one welcomes Charles Pasternak's book *The Molecules within Us*, which can be warmly recommended. Pasternak has the expert knowledge and the communicative skills to lead us safely through some of the most complex science that is currently going on. Even though we may soon

forget the detail of what he tells us, we are left with a memorable and upbeat message: that molecules are what we are, and that disease is mainly due to their malfunction.

This is a book of solid science, packed with a mixture of the latest advances in the molecular biology of the human body, and leavened with the author's opinions on the human condition as most of us experience it. It is a mixture of scientific explanation that an educated lay person can grasp, and of common sense that most will agree with.

As its title implies, *The Molecules within Us* is about the basic units that compose the living cell and all the components that go into making us what we are. There are chapters on nutrition and metabolism, genes and the environment, infection and the immune system, stress (an excellent chapter), the workings of the brain (also very good), and ageing and death. Pasternak should now be prevailed on to write a shorter version of his book that can carry its message to an even wider audience.

Pasternak peppers his account with fascinating anecdotes and curious details, such as this: "In Hyderabad, India, asthmatics are lining up to have a small (2 inches long) live fish known as a murrel. . . dropped down their throats. . . one fish per year, for three successive years, is said to provide a lifelong cure."

Before we give a condescending smile at the naivety of the above, we should bear in mind that we in the West are not immune to embracing alternative treatments such as homeopathy, herbal medicine, aromatherapy and reflexology. I was disappointed that Pasternak did not take a harder line on alternative medicine, but perhaps he was advised not to do so, as it might have alienated some of his readers. Nevertheless, his seventh chapter, "Orthodox versus Alternative Medicine", finally comes down against it. Yet even scientific medicine is finding it hard to defeat the devils of cancer. While some death rates from cancer are declining, such as cervical and testicular cancer, others are increasing, such as liver, prostate and lung cancer in women (although this can be attributed to smoking).

I warmed to Pasternak's message in his chapter on the causes of diseases, when he dismisses the prevailing idea that strenuous exercise will prolong our lives. He concludes that, "provided we have no major health problems, we should just continue to sit in our armchairs, read the scientific journals and newspapers about startling new discoveries . . . and go for a walk to buy fresh fruit and vegetables". How comforting it is to find an expert offering such reassuring advice about the lifestyle that many of us have adopted.

This is an excellent book, possibly a little heavy going for the non-scientist, but not for the readers of *Nature*. □

John Emsley is in the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

Star formation triggered by galaxy collisions

Reinhard Genzel, Dieter Lutz & Linda Tacconi

It is becoming increasingly clear that collisions between galaxies play an important role in galaxy evolution. The ultraluminous infrared galaxies are predominantly powered by enormous star-formation events that are triggered in the last phases of such collisions. These bursts occur just before the galaxies merge to form single elliptical galaxies.

Not until the first two decades of this century was it established that 'galaxies' are large concentrations of stars and interstellar gas, like our Milky Way, but located large distances away. Until the mid-1950s, the accepted view was that these galaxies are 'island universes', very much on their own and having little contact with others. Observations across the entire electromagnetic spectrum, from radio waves to X-rays, have changed this picture significantly. It has become clear that close encounters ('collisions') of galaxies are a common feature in the 'inner city' traffic of the Universe.

The astronomer Fritz Zwicky was the first¹ to call attention to filaments and jets of stars in adjacent galaxies, which he suggested could be large-scale 'tidal effects' due to their mutual gravitational interaction. But this interpretation met with widespread scepticism until the first numerical models showed that the slow but close passages of two galaxies can create long tails, bridges and spiral structures, purely as the result of the gravitational forces and torques between the two colliding galaxies^{2,3}. One representative galaxy pair successfully modelled as such a tidal encounter³ was the NGC 4038/39 system (the 'Antennae'; Fig. 1). Nevertheless, interacting galaxies were still considered as something of a curiosity.

This changed after the mid-1980s as a result of the first space-

borne infrared telescope (IRAS), which surveyed the entire sky in the 12- to 120- μm band (see refs 4–8, for example). Although most of the several tens of thousands of galaxies identified by IRAS are normal galaxies, like our own Milky Way, there are also hundreds of galaxies in the IRAS catalogues whose total energy output is dominated by their emission in the far-infrared band⁸. The most spectacular of these are the ultraluminous infrared galaxies (ULIRGs), which have infrared luminosities between 10^{12} and $10^{13} L_{\odot}$, which is 10^2 to 10^3 times greater than the luminosities of normal galaxies and more akin to extreme active galactic nuclei (AGNs) such as quasars. AGNs are widely believed to be powered by accretion of gas and stars onto massive black holes residing in the centres of the AGN host galaxies (see ref. 9, for example). During the accretion process, a significant fraction of the rest mass energy of the infalling matter is converted into radiation that peaks in the X-ray and ultraviolet bands.

Ground-based follow-up observations of ULIRGs have established that most of them are extreme interacting systems with strong signatures of recent tidal disturbances, in some cases showing that the two galaxies are actually close to merging into a single system¹⁰. Key questions are thus why there is such a strong correlation between infrared luminosity and galaxy interaction, what powers

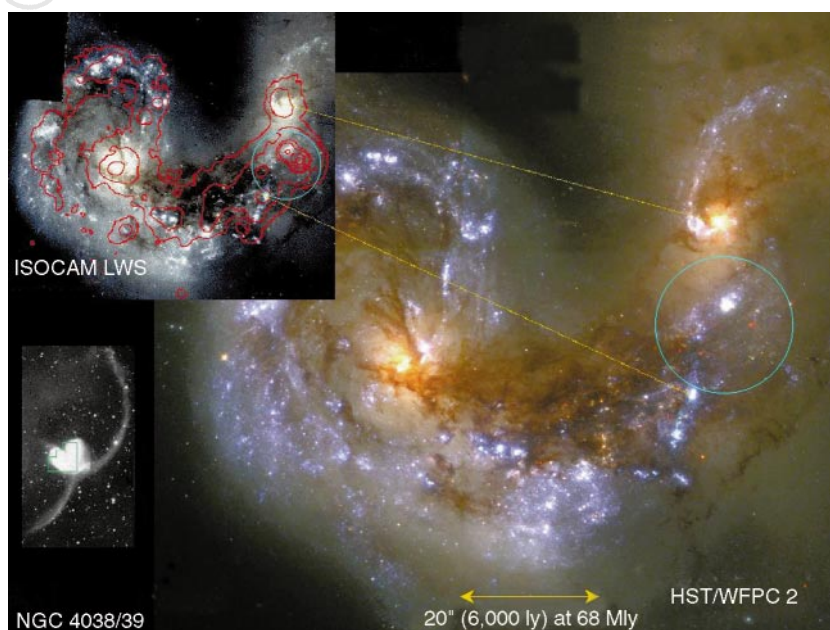


Figure 1 Galaxy collision triggers starburst in the 'Antennae' system (NGC 4038/39). The bottom left inset shows a large-scale image of the two overlapping galaxy disks and the antennae-shaped tidal tails that have given the system its name. The central true colour optical image taken with the WFPC2 camera on board the HST²⁶ shows the two nuclei in orange, as well as a large number of blue (young)

giant star clusters. A turquoise circle marks a dusty (brown/red) region in the overlap region between the two galaxies. Mid-infrared imaging with ISOCAM shows that ~20% of the entire luminosity of the Antennae originates in this compact region²².

these galaxies (active star formation or black holes), and how they are evolving with time.

A case study: Arp220

Recent key observations of the best-studied example of one such spectacular galaxy, the relatively nearby ULIRG Arp220, are summarized in Figs 2 and 3. Hubble Space Telescope (HST) images show that this system is highly disturbed and amorphous. The nuclear region is completely obscured in the visible by a foreground dust lane. In the near-infrared ($\lambda \sim 2 \mu\text{m}$), however, the dust veil becomes partially transparent and a double-peaked structure appears, with a separation of about $0.9''$ ($\sim 1,000$ light years; refs 11, 12). The near-infrared emission comes predominantly from cool supergiant stars with ages between 10 and 30 million years. The double-peaked structure and the northeast to southwest dust lane are also present in rotational line emission of the carbon monoxide (CO) molecule, obtained with millimetre wavelength spatial interferometers^{13,14}. The molecular line emission traces the distribution of the dense, neutral interstellar medium. About $10^{10} M_{\odot}$ of interstellar gas (together with interstellar dust) is concentrated in the nuclear region, comparable to or even exceeding the entire molecular gas content of gas-rich spiral galaxies. Using sensitive Very Long Baseline Interferometry (VLBI), it has been found¹⁵ that a significant fraction of the radio continuum emission in the two peaks comes from a dozen or so unresolved, bright sources of synchrotron emission. The compact radio sources are probably created by recently exploded massive stars (supernovae) that are now ploughing into the surrounding, dense interstellar medium.

Almost all of Arp220's luminosity of $1.5 \times 10^{12} L_{\odot}$ emerges in the

20–200 μm band. The far-infrared radiation comes from interstellar dust grains with a temperature of about 50 K. They efficiently absorb the short-wavelength radiation from stars, or an AGN intrinsically powering the system, and convert it into long-wavelength heat radiation. As a result of this conversion process, however, specific information about the nature of the intrinsic source(s) of luminosity is lost. The best clues about the source(s) powering Arp220 come from spectroscopy of atomic lines and other spectral features. Because of the large central dust obscuration, such spectroscopy is best carried out in the infrared range.

With the launch of the Infrared Space Observatory (ISO) of the European Space Agency (ESA) three years ago, a unique new measuring tool has become available for exploring the infrared spectra of ULIRGs with unprecedented detail and sensitivity. The 2–200- μm band accessible by the two spectrometers on board ISO contains a host of spectral features whose absolute and relative intensities give key information on the character of the intrinsic (ultraviolet) spectral energy distribution of the luminosity source(s). The spectral energy distribution in the ultraviolet band is a key fingerprint distinguishing objects powered by an AGN from those powered by stars. The very hot ($\sim 10^6$ K) accretion disks around black holes in AGNs produce a very hard ultraviolet radiation field, whereas even the most massive stars have much lower temperatures ($\sim 45,000$ K). Measurements with the Short-Wavelength Spectrometer (SWS) in the 2.5–40- μm range show that high excitation atomic lines are prominent in AGNs, requiring hard ultraviolet radiation for their excitation. These same lines are absent or very weak in galaxies powered by stars. As explained in more detail in Box 1, the spectral features observed with ISO in Arp220

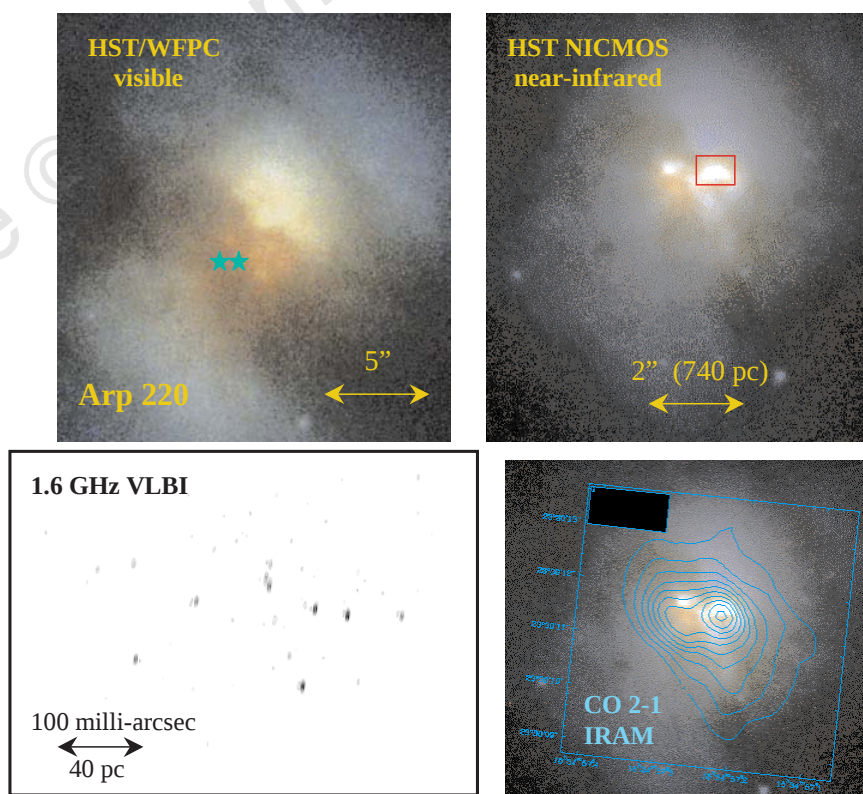


Figure 2 The ultraluminous infrared galaxy Arp220. The HST WFPC2 image in the visible band (top left) shows that the nuclear region of this system (distance 240 million light years) is heavily obscured by a foreground dust lane. In the recent 2- μm NICMOS image (top right), however, the dust veil is partially lifted and the nuclei of two merging galaxies become apparent¹². The positions of the two nuclei are marked as asterisks on the visible image in the top left panel. The

nuclear region is associated with an enormous concentration of dense molecular gas mapped with millimetre interferometry in the CO $J = 2-1$ rotational transition (bottom right: light blue contours superimposed on the NICMOS image¹⁴), and a collection of parsec scale, compact synchrotron radio sources (bottom left; ref. 15). The region encompassed by the radio VLBI map in the bottom left panel is marked by a red rectangle on the top right near-infrared map.

(Fig. 3) and about 60 other ULIRGs provide compelling evidence that the intrinsic ultraviolet radiation field in the nuclear region is soft and resembles that seen in galaxies with active star formation, but not that in AGNs (refs 16–19).

The qualitative picture emerging from these observations is that Arp220 is a fairly advanced merger of two gas-rich spiral galaxies whose enormous luminosity is produced by several million recently formed, massive stars. The near-infrared and radio data indicate that the two emission peaks are probably the two galaxy nuclei that are now at a separation of only about 1,000 light years. As a result of the collision, the interstellar gas of the two galaxies has been streaming towards the minimum of the gravitational potential and is at present forming a ring-like or disk-like configuration. The large concentration of dense gas clouds has triggered a phase of very active star formation (a so-called 'starburst'), with a few hundred new stars forming each year in the central 1,000 light

years. For comparison, three stars are forming each year in the entire Milky Way in a region of more than 50,000 light years across. This phase cannot last very long. The combination of the destructive effects of supernova explosions, seen directly in the radio VLBI map, and the exhaustion of the gas supply will terminate the present level of star-forming activity after only a few tens of millions of years.

Triggering of star formation and AGN activity

The above scenario for what is happening in ULIRGs is supported by numerical simulations. A series of N -body smoothed-particle hydrodynamics (SPH) calculations of colliding galaxies has been carried out (see, for example, refs 20 and 21), including both gaseous and stellar components. As an example, Fig. 4 shows snapshots of the spatial distribution of the young stars and the gas predicted by these models²⁰. According to these calculations, a complete merger of two massive galaxies can occur in a few hundred million years after the first approach, with several consecutive close encounters of ever-decreasing separation. Strong gravitational torques lead to very substantial loss of angular momentum in the interstellar gas component in the first encounter. As a result, the interstellar gas is streaming into the central regions at very high rates and forms high-density concentrations there. Observations of our own and nearby galaxies suggest that the star-formation rate is proportional to the first or second power of mean gas density ($R^* \sim \rho_{\text{gas}}^{1-2}$). If that empirical scaling applies for such colliding systems as well, there should be several brief but very active periods

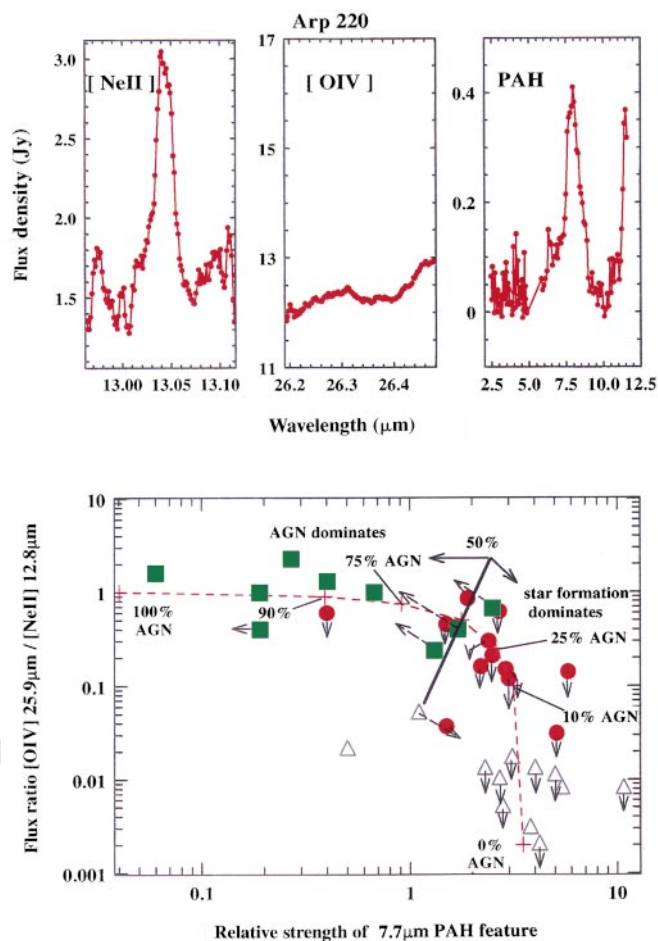


Figure 3 ISO infrared spectra and the new ISO diagnostic diagram. The top three panels show spectra of Arp220 obtained with SWS and ISOPHOT-S. Low-excitation infrared line emission (for example, the 12.8-μm fine structure line from singly ionized neon (Ne II) and emission in the 6 to 11 μm PAH features) are strong, whereas high excitation lines (such as the 25.9-μm line from triply ionized oxygen (O IV)) are faint or absent, all indicative that Arp220 is powered by massive stars, and not an AGN. Bottom: the new ISO diagnostic diagram distinguishes galaxies powered by AGNs and by star formation from their mid-infrared excitation characteristics. Plotted on the vertical axis is the flux ratio of [O IV]/[Ne II] and on the horizontal axis is the flux density ratio of the 7.7 μm PAH feature to the neighbouring continuum. In this diagram, classical AGNs (green squares) and classical starburst galaxies (triangles) are well separated. ULIRGs (red circles) fall in between these two extremes and are composite objects, but are mostly dominated by young massive stars¹⁸.

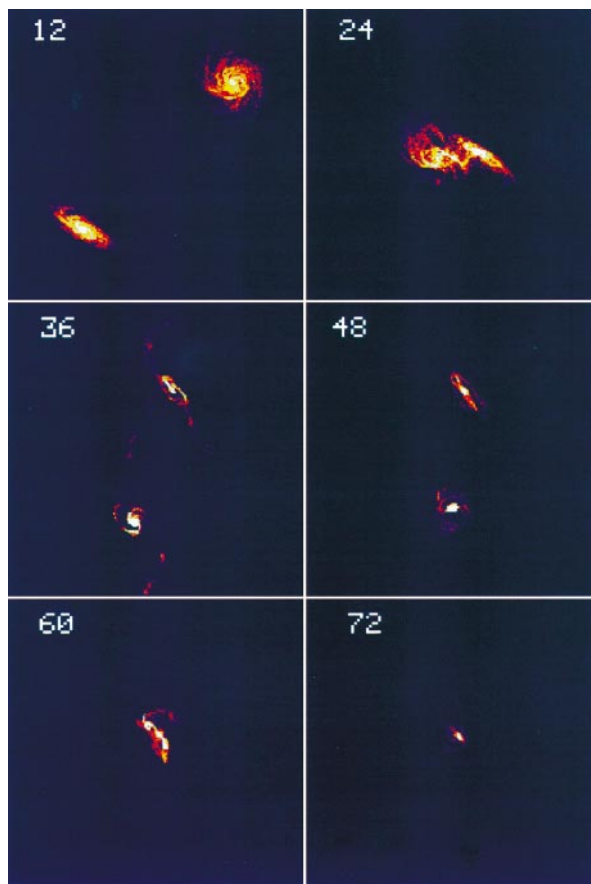


Figure 4 Snapshots of the morphology of the gas and young stellar components (representing distribution and rate of star formation) of a collision of two disk galaxies, as derived from an N -body computer simulation²⁰. Each inset is 230 light years on the side and labelled with the time since the beginning of the collision in the top left panel. One time unit corresponds to 1.3×10^7 years. In the final step after about 1 Gyr (bottom right), the two disk galaxies have merged into a single elliptical galaxy.

Box 1 ISO spectroscopy of ULIRGs

The 2–200 μm band accessible to the two spectrometers on board the Infrared Space Observatory (ISO) contains a host of spectral features whose absolute and relative intensities give key information on the character of the intrinsic spectral energy distribution of the luminosity source(s). Measurements with the Short-Wavelength Spectrometer (SWS) in the 2.5–40 μm range show, for instance, that in AGNs, high-excitation atomic lines, such as quadruply ionized neon, Ne^{4+} (producing characteristic lines at 14 and 24 μm), or triply ionized oxygen, O^{3+} (producing a characteristic line at 25 μm), are prominent. These same lines are completely absent or very weak in galaxies powered by stars. Conversely, AGNs do not show the so called 'PAH' emission features between 3 and 11 μm , whereas the same features are strong in galaxies powered by massive stars. These emission features at 3.3, 6.2, 7.7, 8.7 and 11.2 μm and so on, are probably emitted by large carbon/hydrogen-rich ring molecules (such as polycyclic aromatic hydrocarbons (PAHs)) or very small amorphous carbon clusters that are exposed to moderately intense and soft ultraviolet radiation.

The spectral features observed with ISO in Arp220 (Fig. 3) provide compelling evidence that the intrinsic ultraviolet radiation field in the nuclear region is soft and resembles that seen in galaxies with active star formation, and not that in AGNs^{16–18}. X-ray observations of Arp220 corroborate this conclusion. Data obtained with the ROSAT and ASCA X-ray satellites show that Arp220's emission in the 1 to 10 keV energy band is weaker by a factor of at least a few hundred than in typical AGNs^{24,25}. The only way out is that Arp220 contains an AGN that is hidden even at 10 keV and in the mid- and far-infrared. Although this possibility cannot at present be excluded completely, our results make it very unlikely (see ref. 18, for example).

To investigate whether Arp220 is representative of the entire class of ULIRGs, detailed ISO SWS and ISOPHOT-S spectroscopy has been carried out on a sample of 15 bright ULIRGs¹⁶. In a diagram of flux ratio of high-to-low excitation infrared lines versus strength of the 7.7 μm PAH features, AGNs and starburst galaxies can be cleanly separated (Fig. 3). In this new infrared 'diagnostic diagram', ULIRGs as a class appear to be composite objects with both star formation and AGNs present, but with star formation dominating the overall luminosity in ~80% of the cases¹⁶. Recent additional ISOPHOT-S observations of the 7.7- μm PAH feature of a larger sample of ~60 ULIRGs up to redshift of $z = 0.3$ and luminosity $6 \times 10^{12} L_{\odot}$ further strengthen the statistical significance of this result¹⁹. These data furthermore indicate that the AGN fraction increases for the most luminous galaxies in the sample.

of star formation associated with each close passage of the two galaxy nuclei and peaking in a 'monster' burst in the last encounter just before the final merging²⁰. A small fraction of the gas may even have a low enough angular momentum to make it all the way to the very central regions, fuel the massive black holes there, and trigger AGN activity. This appears to be exactly what is seen in ultraluminous infrared galaxies.

ULIRGs are too distant for individual star-forming regions to be seen directly. In contrast, in the much closer (but also more modest) 'Antennae' galaxy pair discussed above, recent HST and ISO images (Fig. 1) provide compelling proof that a collision between two galaxies leads to fireworks of star formation. The blue spots in the HST image are young star clusters, each containing $\sim 10^3$ to $>10^5$ stars. Whereas these newly formed star clusters are spread throughout the central few thousand light years of the two galaxies, the greatest concentration of present star-forming activity occurs in a highly obscured region in the 'interaction zone' between the two nuclei. This region can be readily identified in mid-infrared images with the camera ISOCAM on board ISO (ref. 22) and is also apparent in millimeter CO line maps.

On the basis of broad-band optical and infrared data and optical and radio spectroscopy, Sanders *et al.*²³ have proposed that the relative importance of AGN activity (relative to starburst activity) increases during the final phases of the merger. The space density of ULIRGs is at least as large as that of quasars of the same luminosity, suggesting that there may in fact be an evolutionary connection between them. In the Sanders *et al.* scenario, the most powerful ULIRGs are predominantly dust-enshrouded AGNs, in a phase just before a classical quasar is revealed. This plausible evolutionary scenario connecting starbursts and AGNs through the time evolution of the merger is only partially supported by the new data. The finding that ULIRGs are often composite objects, with both AGN and starburst activity present in the same system, is in agreement with the Sanders scenario. However, there appears to be no clear trend for advanced mergers (such as Arp220) to be dominated by AGN activity^{18,19}. It is perhaps more likely that the relative importance of starburst and AGN activity is also dependent on the time-dependent, local gas accretion rate onto the central black hole(s) and its radiation efficiency, in addition to the evolutionary phase of the merger¹⁸. High-resolution infrared imaging now underway with the Hubble Space Telescope should test this hypothesis.

Numerical models, as well as observations of the stellar distribution in ULIRGs, indicate that the final product of a merger is a large, dynamically relaxed elliptical galaxy²⁰ (Fig. 4). Most elliptical galaxies and the 'elliptical-like' central bulges of galaxies were formed eight or more billion years ago, in the early, 'galaxy formation' epoch of our Universe. Detailed observations of these very distant objects, now possible with large ground-based telescopes and the Hubble Space Telescope, are expected to provide insight into whether the galaxy-merging processes described here were also important when our Universe was in its infancy. $\square$

Reinhard Genzel, Dieter Lutz and Linda Tacconi are at the Max-Planck Institut für Extraterrestrische Physik, 85740 Garching, Germany.

1. Zwicky, F. Multiple Galaxies. *Ergebnisse der Exakten Naturwissenschaften* **29**, 344–385 (1956).
2. Pfleiderer, J. Gravitationseffekte bei der Begegnung zweier Galaxien. *Z. Astrophys.* **58**, 12–22 (1963).
3. Toomre, A. & Toomre, J. Galactic bridges and tails. *Astrophys. J.* **178**, 623–666 (1972).
4. de Jong, T. *et al.* IRAS observations of Shapley–Ames galaxies. *Astrophys. J.* **278**, L67–L70 (1984).
5. Joseph, R. D., Meikle, W. P. S., Robertson, N. A. & Wright, G. S. Recent star formation in interacting galaxies. I—Evidence from JHK photometry. *Mon. Not. R. Astron. Soc.* **209**, 111–122 (1984).
6. Lonsdale, C. J., Persson, S. E. & Matthews, K. Infrared observations of interacting/merging galaxies. *Astrophys. J.* **287**, 95–107 (1984).
7. Joseph, R. & Wright, G. S. Recent star formation in interacting galaxies. II—Super starburst in merging galaxies. *Mon. Not. R. Astron. Soc.* **214**, 87–95 (1985).
8. Soifer, B. T., Houck, J. R. & Neugebauer, G. The IRAS view of the extragalactic sky. *Annu. Rev. Astron. Astrophys.* **25**, 187–230 (1987).
9. Rees, M. Black hole models of Active Galactic Nuclei. *Annu. Rev. Astron. Astrophys.* **22**, 471–506 (1984).
10. Sanders, D. B. & Mirabel, I. F. Luminous infrared galaxies. *Annu. Rev. Astron. Astrophys.* **34**, 749–792 (1996).
11. Graham, J. R. *et al.* The double nucleus of Arp 220 unveiled. *Astrophys. J.* **354**, L5–L8 (1990).
12. Scoville, N. Z. *et al.* NICMOS imaging of the nuclei of Arp 220. *Astrophys. J.* **492**, L107–L110 (1998).
13. Scoville, N. Z., Yun, M. S. & Bryant, P. Arcsecond imaging of the CO emission in the nucleus of Arp 220. *Astrophys. J.* **484**, 702–719 (1997).
14. Downes, D. & Solomon, P. R. Rotating nuclear disks in ultra-luminous galaxies. *Astrophys. J.* (in the press).
15. Smith, H. E., Lonsdale, C. J., Lonsdale, C. J. & Diamond, P. J. A starburst revealed—luminous radio supernovae in the nuclei of Arp 220. *Astrophys. J.* **493**, L17–L21 (1998).
16. Sturm, E. *et al.* ISO-SWS spectroscopy of Arp 220: a highly obscured starburst galaxy. *Astron. Astrophys.* **315**, L133–L136 (1996).
17. Lutz, D. *et al.* What powers luminous infrared galaxies? *Astron. Astrophys.* **315**, L137–L140 (1996).
18. Genzel, R. *et al.* What powers ultra-luminous IRAS galaxies? *Astrophys. J.* **498**, 579–605 (1998).
19. Lutz, D., Spoon, H. W. W., Rigopoulou, D., Moorwood, A. F. M. & Genzel, R. The nature and evolution of ultraluminous infrared galaxies: A mid-infrared spectroscopic survey. *Astrophys. J.* **505**, L103–L107 (1998).
20. Mihos, J. C. & Hernquist, L. Gas dynamics and starbursts in major mergers. *Astrophys. J.* **464**, 641–663 (1996).
21. Barnes, J. E. & Hernquist, L. Transformation of galaxies. II. Gasdynamics in merging disk galaxies. *Astrophys. J.* **471**, 115–142 (1996).
22. Mirabel, I. F. *et al.* The dark side of star formation in the Antennae galaxies. *Astron. Astrophys.* **333**, L1–L4 (1998).
23. Sanders, D. B. *et al.* Ultraluminous infrared galaxies and the origin of quasars. *Astrophys. J.* **325**, 74–91 (1988).
24. Rigopoulou, D., Lawrence, A. & Rowan-Robinson, M. Multiwavelength energy distributions of ultraluminous IRAS galaxies—I. Submillimetre and X-ray observations. *Mon. Not. R. Astron. Soc.* **278**, 1049–1068 (1996).
25. Kii, T. *et al.* in *X-ray Imaging and Spectroscopy of Hot Plasmas* (eds Makino, F. & Mitsuda, K.) 161 (Univ. Academy Press, Tokyo, 1997).
26. Whitmore, B. NASA-STScI press release STScI-PRC97-34a (1997).

Structural basis for activation of the titin kinase domain during myofibrillogenesis

Olga Mayans*, Peter F. M. van der Ven†, Matthias Wilm‡, Alexander Mues‡, Paul Young‡, Dieter O. Fürst†, Matthias Wilmanns* & Mathias Gautel‡

* European Molecular Biology Laboratory, Hamburg Outstation c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany

† Department for Cell Biology, University of Potsdam, Lenneestrasse 7a, D-14471 Potsdam, Germany

‡ European Molecular Biology Laboratory Heidelberg, Meyerhofstrasse 1, Postfach 10 22 09, D-69012 Heidelberg, Germany

The giant muscle protein titin (connectin) is essential in the temporal and spatial control of the assembly of the highly ordered sarcomeres (contractile units) of striated muscle. Here we present the crystal structure of titin's only catalytic domain, an autoregulated serine kinase (titin kinase). The structure shows how the active site is inhibited by a tyrosine of the kinase domain. We describe a dual mechanism of activation of titin kinase that consists of phosphorylation of this tyrosine and binding of calcium/calmodulin to the regulatory tail. The serine kinase domain of titin is the first known non-arginine-aspartate kinase to be activated by phosphorylation. The phosphorylated tyrosine is not located in the activation segment, as in other kinases, but in the P + 1 loop, indicating that this tyrosine is a binding partner of the titin kinase substrate. Titin kinase phosphorylates the muscle protein telethonin in early differentiating myocytes, indicating that this kinase may act in myofibrillogenesis.

Protein kinases are important in controlling cell proliferation and differentiation and they require specific mechanisms of regulation and substrate recognition¹⁻³. Tight control of the enzymatic activity of protein kinases is achieved by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in carboxy-terminal autoregulatory tails, induced by effector molecules such as Ca²⁺/calmodulin. In all available structures of protein kinases regulated by phosphorylation, one or more phosphorylated residues are always located in the activation segment (Fig. 1). These kinases contain a strictly conserved arginine preceding the catalytic aspartate, and are hence known as RD kinases. The arginine interacts with a phosphorylated residue from the activation segment². In the non-RD kinases of the

myosin-light-chain kinase (MLCK) family, this pattern is not found (Fig. 1), and activation of these kinases by phosphorylation has not been predicted.

The assembly of striated myofibrils in differentiating myocytes involves the controlled integration of hundreds of proteins into the highly ordered macromolecular complex of the sarcomere. The giant protein titin extends over one half of the sarcomeric unit and is crucial in this control^{4,5}. Close to its C terminus, titin contains a MLCK-like kinase domain. Although sequence similarity of the catalytic domain of titin to that of its invertebrate analogue, the serine/threonine kinase domain of twitchin⁶, has indicated that titin and twitchin may have similar function, the differential localization of the kinase domains^{7,8}, their distinct C-terminal regulatory tails^{9,10} and the presence of several titin-specific residues in the active site¹¹

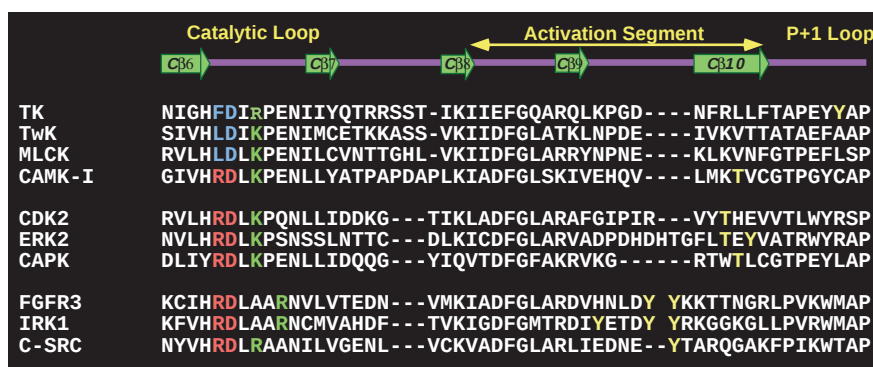


Figure 1 Sequence alignment of active-site regions. Serine/threonine kinases activated by calmodulin or by other calcium-binding proteins: human cardiac titin kinase (TK), twitchin (TwK) from *Aplysia californica*, rat skeletal myosin-light-chain kinase (MLCK), and human calcium/calmodulin-dependent protein kinase-I (CAMK-I); serine/threonine kinases activated by phosphorylation in the activation segment: human cell-division-protein kinase 2 (CDK2), human extracellular-signal-regulated kinase 2 (ERK2), and human cyclic AMP-dependent protein kinase α -catalytic subunit (cAPK); tyrosine kinases: fibroblast growth factor

receptor kinase 3 (FGFR3), insulin-receptor inase (IRK3), and proto-oncogene tyrosine protein kinase c-Src (C-SRC). Yellow, identified phosphorylation sites; red, RD motif; cyan, deviations from the RD motif; green, conserved basic residues (K, R) in the +2 or +4 position relative to the RD motif. The secondary-structure elements shown at the top relate to the three-dimensional structure of titin kinase. The limits of the activation segment and the P + 1 loop are defined by the sequence motifs DFG (EFG in titin kinase) to APE².

hint at a different function for titin. Twitchin is a Ca^{2+} /S100A1-regulated MLCK¹². Despite biochemical evidence for Ca^{2+} /calmodulin binding⁹, however, the activation process and function of titin kinase have remained unknown.

The crystal structure of autoinhibited titin kinase shows how a tyrosine of the P + 1 loop inhibits the active site. The titin kinase structure has provided the basis for determining its dual activation mechanism, which comprises phosphorylation of Y170 and Ca^{2+} /calmodulin binding. Titin kinase is activated in differentiating myocytes, where it phosphorylates the muscle protein telethonin, indicating its probable importance in myofibrillogenesis.

Unusual autoinhibited conformation

The complete autoinhibited form of titin kinase, including the catalytic domain and the regulatory tail (kin1; Fig. 2), could be grown in thin-plate crystals <5 μm thick. We used these crystals to determine the atomic structure of titin kinase at 2.0 Å resolution from synchrotron radiation X-ray data. The overall fold shows the catalytic domain and the autoregulatory C-terminal tail, which wraps around the lower lobe and the active site of the catalytic domain (Fig. 3a, b). The amino-terminal helix of this tail (αR1) is in a similar location to the equivalent helices in twitchin¹⁰ and calcium/calmodulin-dependent kinase-I (CaMK-I)¹³. The C terminus of this tail superimposes with that of twitchin (Fig. 3c) but not with the tail of CaMK-I. The second helix of the tail, αR2 , binds into the ATP-binding site and is followed by a β -strand (βR1) which forms an antiparallel β -sheet with strands βC10 and βC11 in the lower lobe of the catalytic domain. Specific interactions to the catalytic domain are mostly restricted to the two termini, αR1 and βR1 , of the regulatory tail (Fig. 3c). The C-terminal regulatory tail of titin kinase provides the calmodulin-binding site, which mapped to a segment that covers helix αR1 (ref. 9). The surface of this helix exhibits a cluster of basic and hydrophobic residues that is characteristic of Ca^{2+} /calmodulin-binding sequences¹⁴. The analogous segment is also involved in Ca^{2+} /calmodulin binding of other MLCK-like kinases^{13,15,16}.

Unlike other kinases, in which the activation segment undergoes a switch from a 'closed' to an 'open' conformation after serine and/or tyrosine phosphorylation², the activation segment of titin kinase is in an open conformation in the autoinhibited structure (Fig. 4). In the active site, however, the carboxylate group of the invariant catalytic base D127 is involved in a hydrogen-bonding network to R129, Q150 and Y170 (Fig. 4a, d), blocking the active site from access of its protein substrate and thus inhibiting catalysis. In turn,

the hydroxyl group of Y170 is embedded in a network of hydrogen bonds to residues D127, R129 and the main-chain carbonyl group of F126. The important residue in this network is R129, a conserved lysine in all other MLCK-like kinases, which acts as a bridge between D127 and Y170. Steric inhibition of the catalytic base by a tyrosine in its vicinity is reminiscent of the inhibition of the inactive forms of insulin-receptor kinase (IRK)¹⁷ (Fig. 4c) and the MAP kinase ERK2 (ref. 18). These kinases are activated by phosphorylation of at least one tyrosine residue in the activation segment, indicating that tyrosine phosphorylation may be involved in activation of titin kinase. In contrast to ERK2 and IRK, the important inhibitory residue of titin kinase, Y170, is located in the P + 1 loop that connects βC10 and αC4 in the lower lobe of the kinase. This region forms the pocket that accommodates the P + 1 position of the substrate in other kinase peptide-substrate structures^{19,20}.

Tyrosine phosphorylation of titin kinase

To determine whether Y170 is a site of phosphorylation in titin kinase, we made a truncated titin kinase construct (kin4) lacking the regulatory C-terminal tail. Kin4 is predicted to be constitutively active, like C-terminal-truncated twitchin¹⁵. In a yeast two-hybrid analysis, kin4 interacts specifically with overlapping peptides from the regulatory region of titin kinase (Fig. 2b), thus showing the structural integrity of this construct. We introduced a mutation of Y170 (Y170E) into kin4 to abolish the inhibitory tyrosine and, at the same time, to mimic the phosphorylated state¹⁸. The K36A mutation is predicted to disrupt catalytic activity, by analogy to K72A in cAMP-dependent protein kinase²¹. When expressed in C2C12 myocytes, phosphotyrosine was detected in the wild-type kin4 and in the K36A mutant. No phosphotyrosine signal could be detected for the Y170E mutant (Fig. 5a). Similarly, the full-length kinase (kin1) was tyrosine-phosphorylated by extracts from differentiating myocytes, but not by extracts from adult muscle. Again, tyrosine phosphorylation in the mutant kin1(Y170E) was markedly reduced (Fig. 5b). These results indicate that titin kinase is trans-phosphorylated by kinase activities in differentiating muscle at an unusual site in the P + 1 loop, where Y170 is the major phosphorylation site.

Y170 is buried in the autoinhibited, unphosphorylated structure of titin kinase (Fig. 4) like in ERK2 (ref. 18) in which the residues phosphorylated upon activation are equally inaccessible. To make the peptide-binding site accessible, major structural rearrangements are predicted to take place during, or following, tyrosine phosphorylation in both ERK2 (ref. 18) and titin kinase. We immunoprecipitated the native, *in vitro* phosphorylated titin kinase with anti-phosphotyrosine antibodies (Fig. 5c). This indicates that the P + 1 loop may undergo major structural rearrangements involving the exposure to solvent of phosphorylated Y170.

Activated titin kinase phosphorylates telethonin

To study kinase activity of titin kinase, we expressed the truncated titin kinase construct, kin4, in C2C12 myocytes. When we used kin4 in phosphorylation assays with different substrates, we found no myosin light-chain-kinase activity. We detected specifically increased phosphorylation of a protein of relative molecular mass 22,500 (M_r 22K) in day 2 myocyte extracts in the presence of wild-type kin4, but not in the presence of the catalytically inactive kin4 mutant K36A (Fig. 6a). We used mass-spectrometric microsequencing to determine that this band was telethonin, a protein of cardiac and skeletal muscle²². By using phosphorylation assays with recombinant subfragments of telethonin and subsequent sequencing by tandem mass spectrometry, we identified a single phosphorylated serine in the recognition sequence ¹⁵³RRSLS(phospho)RSMSQEAQRG¹⁶⁷, close to the C terminus of telethonin. The identification of telethonin as a sarcomeric Z-disk protein²³ indicates that its phosphorylation may be involved in the control of myofibrillogenesis. This is supported by the observation that C2C12

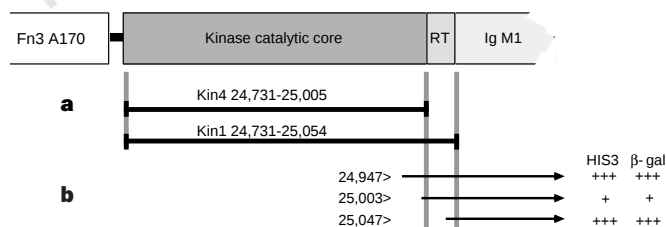


Figure 2 Titin kinase constructs used in this study. The domain pattern of the kinase region of human titin and the residues of the respective constructs are shown. Fn3, fibronectin-III-like domain; Ig, immunoglobulin like domain. **a**, Kin4 lacks the regulatory tail (RT), leaving a constitutively active catalytic core. The kin1 construct includes the regulatory tail. The N-terminal phasing is based on the N termini of DAP kinase³¹ and of *Dictyostellium discoideum* MLCK⁴². **b**, Two-hybrid screens of human skeletal and cardiac complementary DNA libraries with kin4 and kin4(K36A) yield multiple overlapping clones from the regulatory tail, showing that there is a functional binding site in the active site of kin4. The strength of interaction is given by HIS3 signals and by β -galactosidase activity³⁰. The first residue in the cardiac titin sequence of three representative clones is shown at the left.

cells expressing the constitutively active kin4 show a breakdown of normal cytoskeletal architecture (Fig. 6b). In the differentiated myofibril, titin kinase is located $\sim 1 \mu\text{m}$ from the Z-disk, at the edge of the M-band⁸. However, both the titin C terminus and telethonin can be detected by immunofluorescence in dot-like aggregates on stress-fibre-like structures in day 2 myocytes (Fig. 6c). These data indicate that the activation of titin kinase in differentiating myocytes and the resulting phosphorylation of telethonin are involved in the reorganization of the cytoskeleton during myofibrillogenesis. During these events, the titin C terminus

is transiently in close proximity to the titin kinase substrate telethonin.

Activation of titin kinase

To further characterize the mechanism of activation of titin kinase, we used purified, autoregulated kin1 (Fig. 1)⁹. Kin1 showed only weak basal kinase activity towards the substrate telethonin (residues 104–167) (Fig. 7a). Addition of the muscle calcium-binding proteins calmodulin, S100A1₂, CACY or CAPL²⁴ did not increase kinase activity significantly (Fig. 7a). In contrast, the basal activity

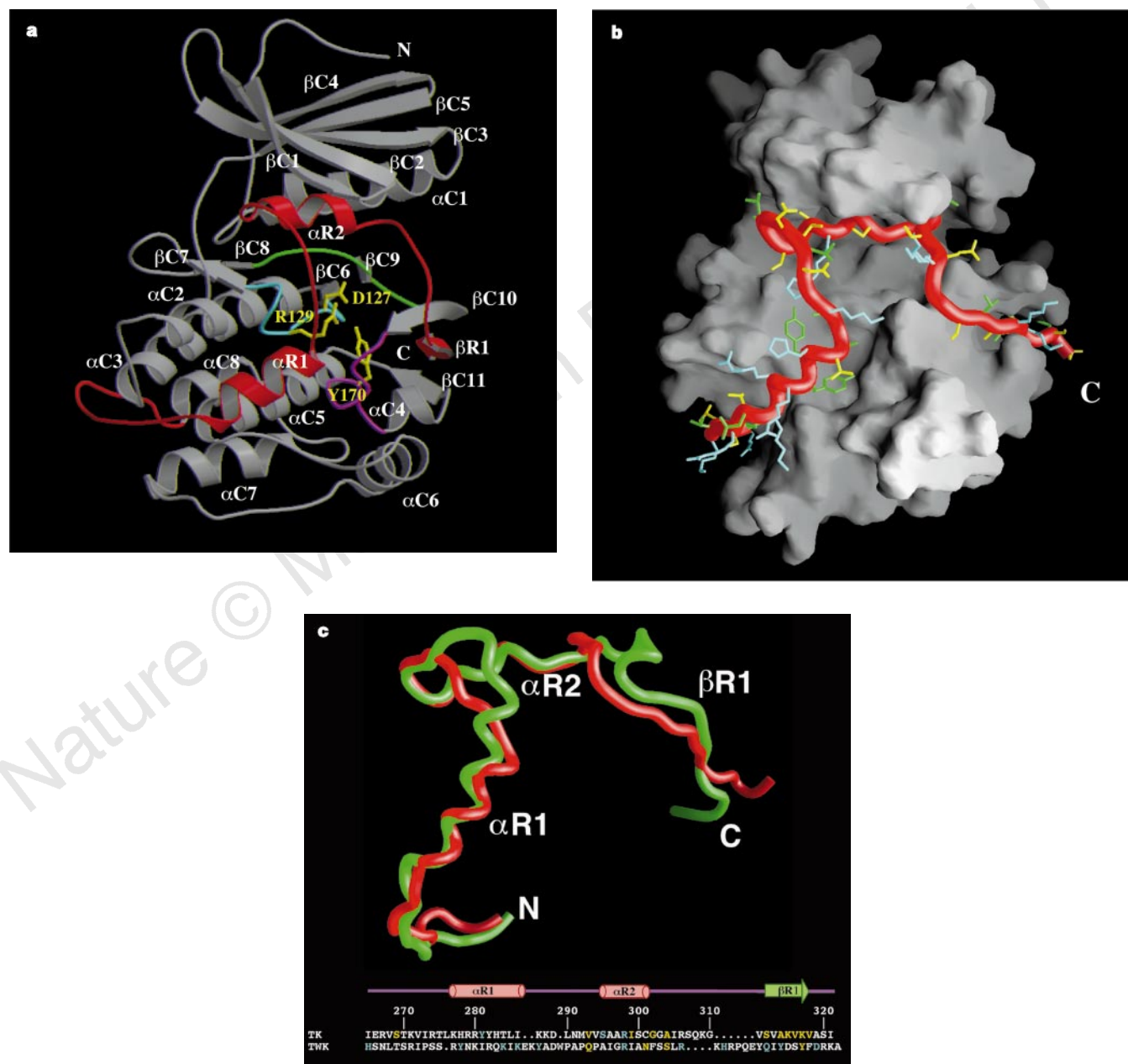


Figure 3 The three-dimensional structure of the autoinhibited form of titin kinase. **a**, Ribbon diagram. Red, regulatory tail; cyan, catalytic loop; green, activation segment; magenta, P + 1 loop. The side chains of residues D127, R129 and Y170 are highlighted in yellow. Y170 is phosphorylated upon activation. The secondary-structural elements are numbered in sequential order. They are labelled with 'C' and 'R', according to their location in the catalytic domain or regulatory tail, respectively. **b**, Surface presentation of the catalytic domain; the regulatory tail is shown as a tube. The side chains of the residues of the regulatory tail are coloured in cyan (basic residues), green (hydrophobic residues) and yellow

(other residues). The central part of the regulatory tail blocks the ATP-binding site. **c**, Top, superposition of the regulatory tails of titin kinase (red) and twitchin¹⁰ (green); bottom, a structure-based sequence alignment using the program ALIGN_FRG⁴³. Residues that are involved in polar interactions to the catalytic domain through their main chains and side chains are coloured in yellow and cyan, respectively. Secondary-structure elements of the regulatory tail of the titin-kinase structure are shown schematically. The residue numbers correspond to the numbering convention used to describe the titin-kinase structure. **a** was prepared with MOLSCRIPT⁴⁴ and **b**, **c** with GRASP⁴⁵.

of the phosphate-mimickry mutant kin1(Y170E) towards telethonin was more than tenfold higher than that of wild-type and was stimulated by up to 100-fold by Ca^{2+} /calmodulin (Fig. 7b). None of the S100 proteins (S100A1, CACY or CAPL) activated the mutant titin kinase (Fig. 7b). These results show that two different autoinhibitory mechanism must be overcome to fully activate titin kinase: the P + 1 loop must be released from the substrate-binding site by tyrosine phosphorylation, mimicked in kin1(Y170E), and the conformation of the C-terminal tail must change as a result of binding of Ca^{2+} /calmodulin. This novel dual autoregulatory mechanism is likely to provide tight control for titin kinase activity during muscle differentiation.

Substrate recognition by titin kinase

In contrast to the RD kinases², the active site of titin kinase does not contain a pocket composed of basic residues that could accommodate a phosphorylated tyrosine. Replacement of the phosphorylated tyrosine of the kinase with substrate, which occurs in IRK^{17,25}, is unlikely because of the catalytic specificity of titin kinase for serine.

None of the other MLCK-like kinases contains a tyrosine in the P + 1 loop (Fig. 1), excluding the possibility of a related activation mechanism. The most significant difference in the phosphorylation site on telethonin, as compared with the sites of phosphorylation of substrates for MLCK and twitchin²⁶ is the presence of an arginine in the P + 1 position. A potential function of the phosphorylated Y170 could, therefore, be to directly bind this arginine, explaining the unusual presence of arginine in the P + 1 position of the titin kinase substrate and the location of the titin kinase phosphorylation site, Y170, in the P + 1 loop. Furthermore, Q150 in the activation segment of titin kinase is replaced by a hydrophobic residue in other MLCK-like kinases (Fig. 1). This glutamine is located in the pocket that accommodates the P + 1 position of the substrate in known kinase peptide-substrate structures²⁷. Therefore, Q150 might be another ligand for the P + 1 arginine of telethonin. The only other known protein kinase that shows the same pattern of a preferred arginine in the substrate P + 1 position and a tyrosine in the kinase P + 1 loop is NIMA¹⁹. The significance of this observation is unknown.

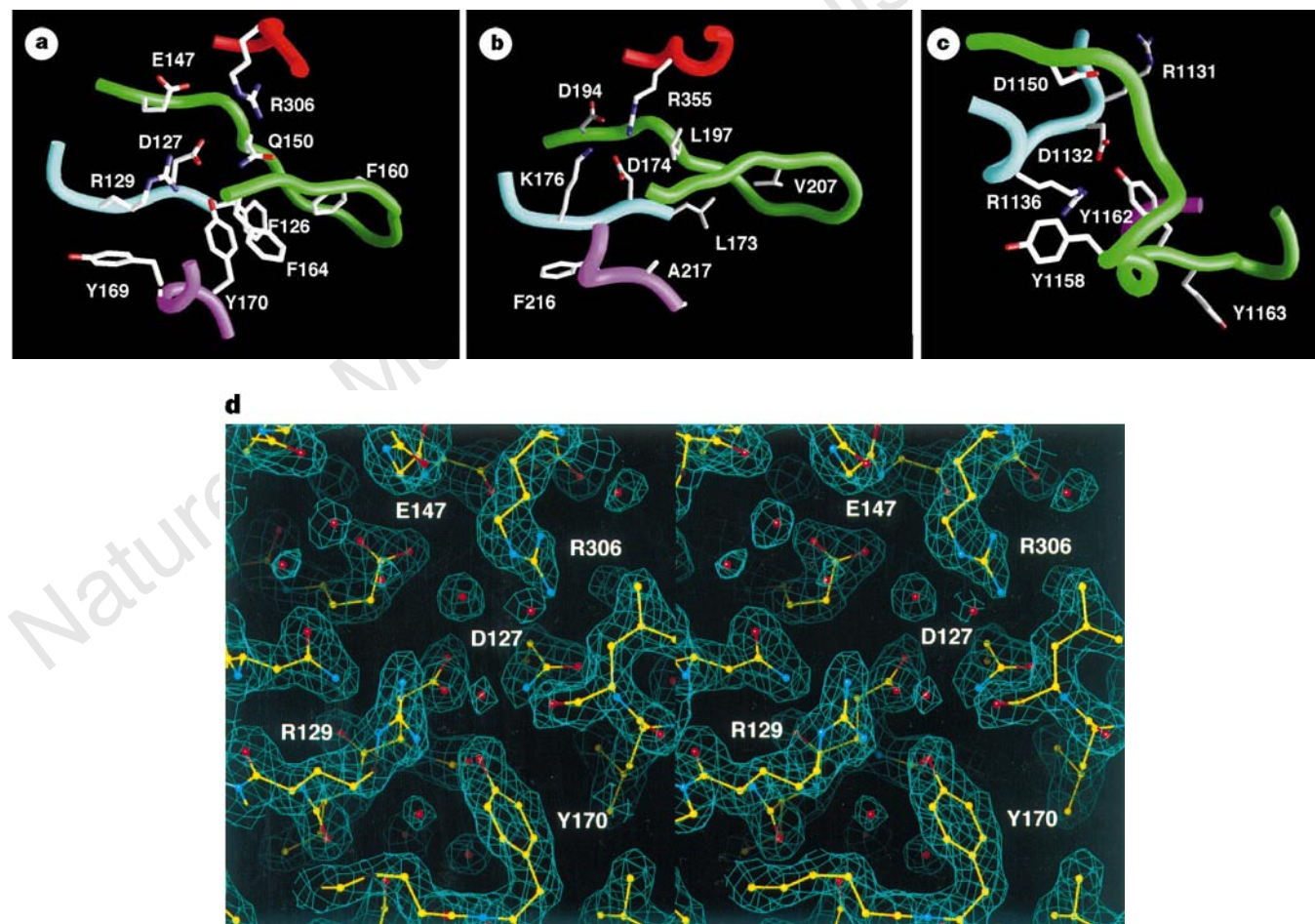


Figure 4 Active-site conformation of the autoinhibited forms of titin kinase, twitchin and IRK. **a**, The active site of titin kinase. The guanidinium group of R129 forms short hydrogen bonds with the side chains of D127 and Y170 from the P + 1 loop. There is a weak direct hydrogen bond between D127 and Y170 (3.1 Å in length). D127 forms further hydrogen bonds with Q150. **b**, Twitchin active site¹⁰. The catalytic aspartate, D174 forms hydrogen bonds with K176, Q200 and R355 from the regulatory tail. At the position of Y170 in titin kinase, there is an alanine in twitchin. In the autoinhibited twitchin structure¹⁰, the catalytic aspartate is blocked by a salt bridge with an arginine (R355 in twitchin) of the regulatory tail, suggesting a different activation mechanism than for titin kinase. In titin kinase, the equivalent

arginine, R306, does not interact with the catalytic aspartate. **c**, Active site of the autoinhibited form of IRK¹⁷. The catalytic aspartate, D1132, is bound to Y1162. This bond is disrupted after phosphorylation of Y1162, accompanied by phosphorylation of two other tyrosines and induces a conformational change of the activation segment from a closed to an open conformation²⁵. The colour codes of the tubes are as in Fig. 3a. **d**, Stereo view of a $2F_o - F_c$ electron-density map, using phases of the final model, contoured at 1.3σ . The electron density shown covers several active-site residues and solvent molecules. Some titin-kinase residues are labelled. **a–c** were prepared with GRASP⁴⁵ and **d** with program O (ref. 46).

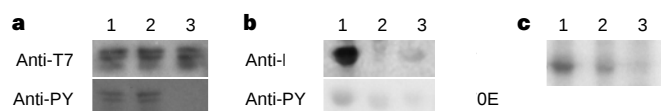


Figure 5 Titin kinase is tyrosine-phosphorylated on Y170 in differentiating C2C12 cells. **a**, Transfected in4 (lane 1), kin4(K36A) (lane 2) and kin4(Y170E) (lane 3) were immunoprecipitated with a rabbit antibody against titin kinase (α -TK-ra; ref. 9) and phosphotyrosine was detected on blots of the immunoprecipitates with the monoclonal antibody 4G10 (anti-PY). The immunoprecipitated kinase was detected with the anti-T7 tag antibody (anti-T7). The absence of a phosphotyrosine signal in kin4(Y170E) indicates that Y170 is the major tyrosine phosphorylation site. The phosphorylation of the catalytically inactive kin4(K36A) indicates that titin kinase is transphosphorylated by upstream kinase activities rather than intramolecularly autophosphorylated. **b**, Similarly, the recombinant full-length kinase kin1 (WT) is markedly tyrosine-phosphorylated *in vitro* by cytosolic extracts from differentiating day 2 C2C12 cells (lane 1) but only weakly from adult psoas muscle extracts (lane 3) as visualized in blots with 4G10. The phosphotyrosine signal of the kin1(Y170E) mutant is markedly reduced (lane 1, bottom), indicating that Y170 is the major phosphorylation site in the autoinhibited form of titin kinase also. Lanes 2 and 3, bottom: control without cell extracts; 1 μ g of enzyme was loaded per lane. **c**, the phosphorylated tyrosine in titin kinase is accessible to anti-phosphotyrosine antibodies in the native state of the enzyme, indicating that it is exposed to solvent and that major structural rearrangements of the P + 1 loop occur in the activated kinase. Kin1 was phosphorylated as in **a** in the presence of [γ - 32 P]ATP (lane 1) and immunoprecipitated with the 4G10 antibody. The autoradiograph shows the presence of phosphorylated titin kinase in the immunoprecipitated fraction (lane 2) and its near absence in the unbound supernatant (lane 3).

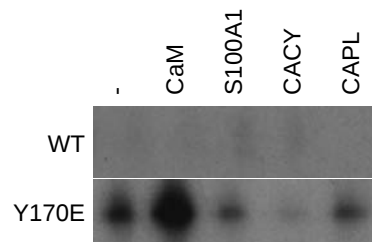


Figure 7 Full activation of titin kinase requires both tyrosine phosphorylation and Ca^{2+} /calmodulin. Wild-type (WT) kin1 shows very low basal activity towards telethonin(104–167) (lane 1); this activity is not stimulated by Ca^{2+} /calmodulin (CaM), Ca^{2+} /S100A1 (S100), Ca^{2+} /CACY (CACY), or Ca^{2+} /CAPL (CAPL). In these assays, we used 0.1 μ g purified enzyme. Autoradiographs were exposed for 12 h. The phosphate-mimicry mutant kin1(Y170E) shows elevated basal activity (lane 1) which is stimulated markedly by Ca^{2+} /calmodulin. The addition of Ca^{2+} /S100A1, Ca^{2+} /CACY or Ca^{2+} /CAPL has no activating effect; the presence of Ca^{2+} /CACY suppresses basal titin-kinase activity by about tenfold, indicating that S100 proteins indeed modulate the activity of titin kinase. The addition of 10 μ M Zn^{2+} in the assays using S100 proteins did not increase kinase stimulation but inhibited kin1(Y170E) as it inhibited *C. elegans* twitchin²⁶.

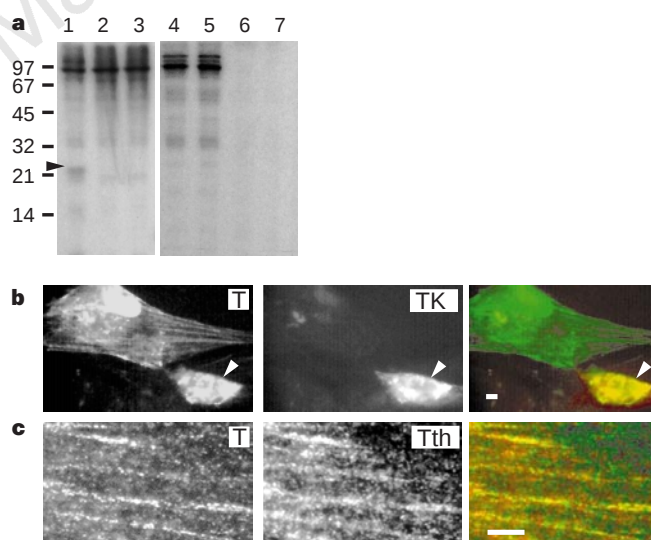


Figure 6 Titin kinase phosphorylates the muscle protein telethonin. **a**, Phosphorylation of muscle proteins by constitutively active kin4 was done using tyrosine-phosphorylated enzyme from transfected C2C12 cells. To suppress background calcium-activated kinases in the myocyte extracts, EGTA was used. In extracts of day 2 myocytes, increased phosphorylation of a protein of $M_r \sim 22$ K occurs in the presence of kin4 (lane 1). In the presence of kin4(K36A) (lane 2) or in blank assays (lane 3), phosphorylation of this band is not increased. Kin4 shows no discernible activity towards proteins in adult psoas cytosol (lane 4) or myofibrils (lane 6) compared with controls (lanes 5 and 7), indicating that the activity of titin kinase is directed towards proteins in differentiating myocytes. The constitutively active enzyme has no myosin-light-chain kinase activity (lane 6).

b, the presence of constitutively active kin4 in myogenic cells disrupts myofibrillogenesis. Non-transfected cells at day 2 show the typical alignment of titin (top panel, visualized by anti-Z1Z2-ra staining³⁰) along stress-fibre-like structures. In the transfected cells (arrowhead in all three panels), the alignment of titin along the actin cytoskeleton is disrupted and titin (green) and kin4 (red, visualized with the anti-T7-tag antibody in the bottom panel) are randomly distributed. **c**, The C terminus of titin loops back onto the stress-fibre-associated N-terminal portion, and localizes with telethonin. In day 2 myocytes, C-terminal epitopes of titin (visualized with the anti-titin antibody T30 (ref. 47) localize with telethonin on stress-fibre-like structures in dot-like aggregates. T, Titin; Tth, telethonin; TK, transfected titin kinase. Scale bars: 10 μ m.

Discussion

The titin kinase structure exhibits dual inhibition of the active site: the catalytic aspartate is blocked by Y170 from the P + 1 loop, and the ATP-binding site is blocked by helix α R2 from the regulatory tail. Inhibition is removed by a new, dual-activation process, involving phosphorylation of Y170 by an unknown kinase and potential co-factors and the binding of Ca^{2+} /calmodulin. Titin kinase is, to our knowledge, the only kinase known to be activated by phosphorylation of a tyrosine from the P + 1 loop and to lack the RD motif. We propose that, on activation of titin kinase, the central part of the regulatory tail is expelled from the active site whereas the flanking interactions of α R1 and β R1 remain in contact with the catalytic domain. It remains to be determined whether activation by phosphorylation in the P + 1 loop will effect substrate specificity as a general mechanism for kinase regulation, as in protein kinase C²⁸. Although all members of the MLCK family bind Ca^{2+} /calmodulin, other calcium-binding proteins are involved in activation; for example, twitchin is activated by the dimeric S100A1 protein^{12,16}. It will be important to determine the specificity and complementarity of different calcium-binding proteins in the activation of these protein kinases, which appears to be a multistep process in most cases. □

Methods

Plasmids and two-hybrid screens. The titin kinase fragments kin1 (EMBL accession number X90568; amino acids (aa) 24,731–25,054) and kin4 (aa 24,731–25,005) and their mutants were cloned into a modified pCMV5 vector²⁹ with an N-terminal T7-tag (Invitrogen) sequence (MTGGQQMGK). The N-terminal phasing was based on MLCK-like kinases without extra N-terminal domains, for reasons of expression stability. For two-hybrid screens, we inserted kin4 and its mutants into a modified pLexA plasmid³⁰. Two-hybrid screens and analysis were done as described³⁰. All cloning and mutagenesis steps followed standard procedures.

Cell culture. We cultured C2C12 cells in DMEM, 20% fetal calf serum and 4.5% glucose at 37 °C. Differentiation was induced by moving cells to low-serum medium (DMEM, 4% horse serum). Transfection with lipofectamine (Gibco-BRL, UK) followed standard procedures. For immunoprecipitation and kinase assays, day 1 cells from 10-cm dishes were lysed in 200 μ l 20 mM HEPES, pH 7.2, 5 mM MgCl_2 , 1 mM EGTA, 1 mM dithiothreitol (DTT), 1 mM Na_3VO_4 , 0.1% Triton X-100 and a cocktail of protease inhibitors (Boehringer).

Antibodies and immunochemistry. An affinity-purified polyclonal rabbit antibody was used against telethonin residues 104–167 (ref. 23). Western blots were performed by standard procedures using the enhanced chemoluminescence (ECL) kit (Amersham). Anti-phosphotyrosine blots were done with the monoclonal antibodies 4G10 (Upstate Biotechnology). Immunoprecipitation was carried out on protein-A beads using the anti-T7-tag (Novagen) or 4G10 monoclonal antibodies, or the anti-titin-kinase polyclonal antibody⁹ (see below) using standard protocols.

Phosphorylation assays. Kinase assays from immunoprecipitates were done as described³¹ in assay buffer (20 mM HEPES, pH 7.2, 5 mM MgCl_2 , 0.5 mM CaCl_2 , 1 mM DTT, 0.2 mM ATP and 1 μ Ci [γ -³²P]ATP, 3,000 Ci mM^{-1}). Assays with purified kinases and telethonin were carried out with 1 ng to 0.4 μ g enzyme in 20 μ l kinase assay buffer, 20 μ g ml^{-1} calmodulin or S100 proteins, and 2 μ g substrate protein. Assays were stopped after 20 min at 30 °C by addition of SDS sample buffer³² and analysed on 15–18% SDS–polyacrylamide gels³². Tyrosine kinase assays were done with 100 μ g ml^{-1} kin1 and 1 μ l cell lysate supernatant in 20 mM MOPS, pH 7, 5 mM MnCl_2 , 1 mM Na_3VO_4 , 0.1 mM ATP and 2 μ Ci [γ -³²P]ATP, 3,000 Ci mM^{-1} , as described³³. Dried gels were autoradiographed for 2–12 h. Quantification of calmodulin- and S100-induced activation and basal activity was done as described²⁶. Phosphoproteins were enriched for microsequencing by chelating chromatography²⁷ and by SDS–PAGE.

Mass spectrometry. The phosphorylated protein was purified on a self-assembled 100-nl Poros^a R2 column (Perceptive Biosystems) and analysed on a triple-quadrupole mass spectrometer (API III, PE-Sciex)³⁴. The protein was digested in 0.1 M NH_4HCO_3 by adding 0.1 μ g trypsin. Half of the resulting peptide mixture was purified on a 100-nl Poros^a R3 column into a nano

electrospray capillary. The sample was analysed on the triple-quadrupole mass spectrometer in negative-ion mode. A phosphorylated peptide was detected and its mass measured using a precursor ion scan of the PO_3^- ion (0.079K)³⁵. The second half of the peptide mixture was desalted as described³⁴, but eluted in 60% methanol, 35% water and 5% formic acid and analysed in positive-ion mode. The doubly charged peptide ion ($M + 2H$)²⁺ was selected for fragmentation in the mass spectrometer. Its sequence and the phosphorylation site could be deduced from the fragment spectrum.

Expression and purification. Human kin1 or the mutant kin1(Y170E), which includes the catalytic domain and the C-terminal regulatory extension (EMBL accession number X90568; aa 24,731–25,054; for simplicity, the residue numbering has been changed by assigning the first residue of the construct as '1'), was purified to homogeneity from sf9 cells infected with recombinant baculoviruses. Purified titin kinase was dialysed into 20 mM Tris–HCl, pH 7.4, 50 mM NaCl, 1 mM EDTA, 1 mM DTT and 1 mM NaN_3 and concentrated in a Centricon-10 to about 4–7 mg ml^{-1} for crystallization. Telethonin fragments were cloned from a primary pGAD10 clone into an N-terminally His₆-tagged PET3d vector. Expression of His₆-fusion proteins followed standard procedures. Recombinant calmodulin, S100A1₂, CACY (calcyclin) and CAPL were purified from *Escherichia coli* on phenyl sepharose and anion-exchange chromatography as described^{24,35}. For some assays, S100A1₂ from bovine brain was used (Sigma) with no differences to the recombinant protein. Purity of proteins was assessed by SDS–PAGE and mass spectrometry.

Crystallization and X-ray structure determination. Crystals were obtained from 1.1 M sodium/potassium tartrate, 2.5% (v/v) ethanol, 25 mM sodium acetate, pH 4.9, and 25 mM imidazole, pH 7.5, using the vapour-diffusion method. The crystals grew as thin plates with dimensions of 5 μ m $\times$ 10 μ m $\times$ 300 μ m. The thickness of these crystals was slightly improved by macroseeding and the use of oil layers over the reservoir as described³⁶. Crystals were shock-frozen at 100 K using 12% glycerol as cryoprotectant. A native X-ray data set was collected up to 2.0 Å resolution on the wiggler beamline BW7B (EMBL Hamburg Outstation) using a wavelength of 0.8373 Å. The data were recorded in a high-resolution sweep (132 frames, frame width 0.8 degrees, crystal-to-detector distance 180 mm) and a low-resolution sweep (83 frames, frame width 1.2 degrees, detector distance 340 mm) on a 345-mm Mar imaging plate scanner. We used the DENZO and SCALEPACK software³⁷ for processing and reduction of data, which consisted of 53,151 unique reflections with a redundancy of 3.7, a completeness of 96% and an $R_{\text{sym}}(\text{I})$ of 7.7% overall (21,052 unique reflections in the highest resolution shell, 2.39–2.0 Å, with a multiplicity of 2.6, a completeness of 94.5% and an $R_{\text{sym}}(\text{I})$ of 23.9%). The crystals belong to the $P2_12_12_1$ space group with cell dimensions of $a = 78.6$ Å, $b = 89.9$ Å, $c = 113.3$ Å and two molecules per asymmetric unit. The non-crystallographic two-fold axis is parallel to the crystallographic b axis at $x = 0.5$ and $z = 0.09$. Initial phases were obtained from a trimmed model of the catalytic domain of twitchin¹⁰ using the molecular-replacement software package AMoRe³⁸. A Sigma-A³⁹ weighted, $2F_o - F_c$ electron-density map calculated with the initial molecular-replacement phases was used for NCS averaging using the AVE software, with mask creation and manipulation done in MAMA⁴⁰. We used the molecular dynamics slow-cooling protocol of X-PLOR⁴¹ for structure refinement. We applied bulk solvent correction, overall anisotropic B -factor scaling, restrained NCS and restrained individual B -factor refinement. The final model includes 5,206 protein non-hydrogen atoms and 514 solvent atoms for the two copies in the asymmetric unit. The overall R -factor is 20.7% and R_{free} is 24.8% for all observed data between 40 and 2.0 Å resolution (in the outer-resolution shell, 2.28–2.0 Å, these values are 22.4% and 27.4%, respectively). The test data set corresponds to 5% of the total data and was generated with FREERFLAG³⁹.

Received 27 May; accepted 14 September 1998.

1. Taylor, S. S., Radzio-Andelzelm, E. & Hunter, T. How do protein kinases discriminate between serine/threonine and tyrosine? Structural insights from the insulin receptor tyrosine kinase. *FASEB J.* **9**, 1255–1266 (1995).
2. Johnson, L. N., Noble, M. E. M. & Owen, D. J. Active and inactive protein kinases: structural basis for regulation. *Cell* **85**, 149–158 (1996).
3. Johnson, L. N., Lowe, E. D., Noble, M. E. M. & Owen, D. J. The structural basis for substrate recognition and control by protein kinases. *FEBS Lett.* **430**, 1–11 (1998).
4. Trinick, J. Titin as a scaffold and spring. *Curr. Biol.* **6**, 258–260 (1996).
5. Maruyama, K. Connectin/titin, giant elastic protein of muscle. *FASEB J.* **11**, 341–345 (1997).
6. Heierhorst, J. et al. Autophosphorylation of molluscan twitchin and interaction of its kinase domain with calcium/calmodulin. *J. Biol. Chem.* **269**, 21086–21093 (1994).

7. Vibert, P., Edelstein, S. M., Castellani, L. & Elliot, B. W. Mini-titins in striated and smooth molluscan muscles: structure, location and immunological crossreactivity. *J. Muscle Res. Cell Motil.* **14**, 598–607 (1993).
8. Obermann, W. M. J. *et al.* The structure of the sarcomeric M band: localization of defined domains of myomesin, M-protein and the 250 kD carboxy-terminal region of titin by immunoelectron microscopy. *J. Cell Biol.* **134**, 1441–1453 (1996).
9. Gautel, M. *et al.* A calmodulin-binding sequence in the C-terminus of human cardiac titin kinase. *Eur. J. Biochem.* **230**, 752–759 (1995).
10. Kobe, B. *et al.* Giant protein kinases: domain interactions and structural basis of autoregulation. *EMBO J.* **15**, 6810–6821 (1996).
11. Sebestyén, M. G., Fritz, J. D., Wolff, J. A. & Greaser, M. L. Primary structure of the kinase domain region of rabbit skeletal and cardiac titin. *J. Muscle Res. Cell Motil.* **17**, 343–348 (1996).
12. Heierhorst, J. *et al.* Ca²⁺/S100 regulation of giant protein kinases. *Nature* **380**, 636–639 (1996).
13. Goldberg, J., Nairn, A. C. & Kuriyan, J. Structural basis for the auto-inhibition of calcium/calmodulin-dependent protein kinase I. *Cell* **84**, 875–887 (1996).
14. Crivici, A. & Ikura, M. Modular and structural basis of target recognition by calmodulin. *Annu. Rev. Biomol. Struct.* **24**, 85–116 (1995).
15. Heierhorst, J. *et al.* Phosphorylation of myosin regulatory light chains by the molluscan twitchin kinase. *Eur. J. Biochem.* **233**, 426–431 (1995).
16. Chin, D., Winkler, K. E. & Means, A. R. Characterisation of substrate phosphorylation and use of calmodulin mutants to address implications from the enzyme crystal structure of calmodulin-dependent protein kinase I. *J. Biol. Chem.* **272**, 31235–31240 (1997).
17. Hubbard, S. R. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO J.* **16**, 5572–5581 (1997).
18. Zhang, F. *et al.* Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. *Nature* **367**, 704–711 (1994).
19. Songyang, Z. *et al.* A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinase I and II, phosphorylase kinase, calmodulin-dependent kinase II, CDK5 and Erk1. *Mol. Cell. Biol.* **16**, 6486–6493 (1996).
20. Lowe, E. D. *et al.* The crystal structure of a phosphorylase inase peptide substrate complex: kinase substrate recognition. *EMBO J.* **16**, 6646–6658 (1997).
21. Taylor, S. S. *et al.* A template for the protein kinase family. *Trends Biochem. Sci.* **18**, 84–89 (1993).
22. Valle, G. *et al.* Telethonin, a novel sarcomeric protein of heart and skeletal muscle. *FEBS Lett.* **415**, 163–168 (1997).
23. Mues, A. *et al.* Two immunoglobulin-like domains of the Z-disk portion of titin interact in a conformation-dependent way with telethonin. *FEBS Lett.* **428**, 111–114 (1998).
24. Engelkamp, D., Schäfer, B. W., Erne, P. & Heizmann, C. W. S100 alpha, CAPL, and CACY: molecular cloning and expression analysis of three calcium-binding proteins from human heart. *Biochemistry* **31**, 10258–10264 (1992).
25. Hubbard, S. R., Wei, L., Ellis, L. & Hendrickson, W. A. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372**, 746–754 (1994).
26. Heierhorst, J. *et al.* Substrate specificity and inhibitor sensitivity of Ca²⁺/S100-dependent protein kinases. *Eur. J. Biochem.* **242**, 454–459 (1996).
27. Songyang, Z. *et al.* Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr. Biol.* **4**, 973–982 (1994).
28. Konishi, H. *et al.* Activation of protein kinase C by tyrosine phosphorylation in response to H₂O₂. *Proc. Natl Acad. Sci. USA* **94**, 11233–11237 (1997).
29. Andersson, S. *et al.* Cloning, structure, and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. *J. Biol. Chem.* **264**, 8222–8229 (1989).
30. Young, P., Ferguson, C., Bañuelos, S. & Gautel, M. Molecular structure of the sarcomeric Z-disk: two types of titin interactions lead to an asymmetrical sorting of α-actinin. *EMBO J.* **17**, 1614–1624 (1998).
31. Cohen, O., Feinstein, E. & Kimchi, A. DAP-kinase is a Ca²⁺/calmodulin-dependent, cytoskeletal-associated protein kinase, with cell-death inducing functions that depend on its catalytic activity. *EMBO J.* **16**, 998–1008 (1997).
32. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
33. Weijland, A. *et al.* The purification and characterization of the catalytic domain of Src expressed in *Schizosaccharomyces pombe*. *Eur. J. Biochem.* **240**, 756–764 (1996).
34. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**, 1–8 (1996).
35. Dedman, J. R. & Kaetzel, M. A. Calmodulin purification and fluorescence labeling. *Methods Enzymol.* **102**, 1–8 (1983).
36. Chayen, N. E. A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *J. Appl. Crystallogr.* **30**, 198–202 (1997).
37. Otkinowski, Z. in *Proceedings of the CCP4 Study Weekend* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Daresbury Lab., 1993).
38. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 577–587 (1994).
39. Collaborative Computational Project 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–767 (1994).
40. Kleywegt, G. T. & Jones, T. A. in *Proceedings of the CCP4 Study Weekend* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC, Daresbury Lab., 1994).
41. Brünger, A. T., Kuriyan, J. & Karplus, M. Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–466 (1987).
42. Tan, J. L. & Spudich, J. Characterization and bacterial expression of the *Dictyostelium* myosin light chain kinase cDNA. *J. Biol. Chem.* **266**, 16044–16049 (1991).
43. Cohen, G. E. A program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallogr.* **30**, 1160–1161 (1997).
44. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallogr.* **24**, 251–259 (1991).
45. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insight from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
46. Jones, T. A. Diffraction methods for biological macromolecules. Interactive computer graphics: FRODO. *Methods Enzymol.* **115**, 157–171 (1985).
47. Fürst, D. O., Osborn, M., Nave, R. & Weber, K. The organization of titin filaments in the half-sarcomere revealed by monoclonal antibodies in immunoelectron microscopy: a map of ten non-repetitive epitopes starting at the Z line extends close to the M line. *J. Cell Biol.* **106**, 1563–1572 (1988).

Acknowledgements. We thank M. Saraste for support in the early stages of this project and J. Heierhorst for stimulating and fruitful discussions.

Correspondence and requests for materials should be addressed to M.W. (e-mail: Wilmanns@embl-hamburg.de) or M.G. (e-mail: Gautel@embl-heidelberg.de). Coordinates and structure factors have been deposited at the Protein Data Base (accession number 1TKI).

Carbon monoxide emission as a precise tracer of molecular gas in the Andromeda galaxy

N. Neininger^{*†‡}, M. Guélin[†], H. Ungerechts[§], R. Lucas[†] & R. Wielebinski[‡]

^{*} Radioastronomisches Institut der Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany

[†] Institut de Radioastronomie Millimétrique, 300, rue de la piscine, F-38406 St Martin d'Hères, France

[‡] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany

[§] Instituto de Radioastronomía Milimétrica, Avenida Divina Pastora 7, E-18012 Granada, Spain

Stars are known to form in clouds of cold molecular hydrogen, which are relatively poorly understood despite being one of the main components of the interstellar medium. The problem is that H_2 is invisible in the cold interstellar medium, so its distribution and motion must be inferred from observations of minor constituents of the clouds, such as carbon monoxide and dust. Most of our present knowledge comes from observations of CO emission, but there is much debate on whether this is an effective tracer of H_2 : it might miss a large fraction of the molecular gas¹. It is difficult to address this question on the basis of observations

within the Milky Way alone, whose edge-on orientation makes it hard to discern the distant cloud structures. We have therefore surveyed the CO emission of the molecular clouds of M31 (the Andromeda galaxy), the nearest spiral galaxy to the Milky Way, and investigated the extent to which it follows the extinction of starlight by dust. We find a remarkably tight association between the CO emission and the dust, from which we conclude that CO does indeed trace all of the molecular gas.

M31 is not only nearby, but its distance is exceptionally well known (2.5 ± 0.1 million light-years², or 0.78 Mpc) and it is inclined by 12° to the line of sight, enough to reveal a magnificent pattern of stars and dust lanes³. This makes it an ideal object for detailed studies of large structures. Whereas the stellar and H I content of M31 has been extensively studied^{3–5}, little has been known concerning the molecular component. The only complete CO millimetre-line survey⁶ has a poor angular resolution ($8.7'$, or 2.0 kpc along the major axis) and even the FCRAO survey in progress⁷ is too coarse to resolve the spiral arms. Our new CO map (Fig. 1) covers one-third of M31's disk with a $23''$ beam (90 pc), small enough to resolve Orion-like cloud complexes. We have mapped in addition the brightest complexes with a $2''$ beam, thus resolving individual clouds (Fig. 2).

The most striking result is the correlation between CO and the dark lanes visible on the optical image (Fig. 1b). It is often argued¹ that CO emission does not trace the molecular gas properly, for two reasons. First, CO molecules are more easily photodissociated than H_2 and could be destroyed in diffuse H_2 clouds. Second, the CO $J = 1 \rightarrow 0$ line is optically thick and its emission could be biased toward warm regions. Allen and co-workers^{8,9}, for example, argue

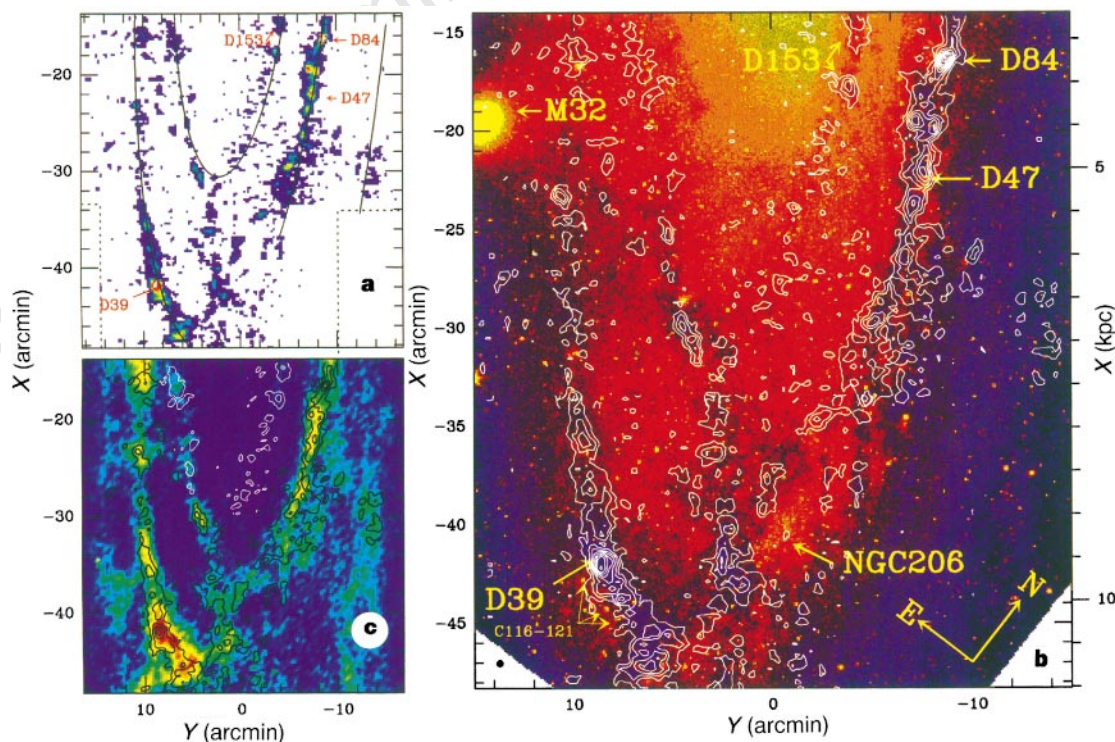


Figure 1 CO line intensity distribution. **a**, The velocity-integrated CO intensity, $I^*(CO) = \int T_A^*(CO_{1-0}) dv$, in the southern half of M31. Observations were made 'on-the-fly' with the IRAM 30-m telescope by using a basket-weaving procedure: the telescope was moved back and forth at a speed of $4'' s^{-1}$ in the Y direction with steps in X of $8''$, then back and forth in the X direction, with steps of $8''$ in Y. The data, recorded every 1 s, were reduced by using the de-stripping algorithm of Emerson and Gräve²¹ adapted by P. Hoernes to spectral-line observations²². In the process, the resolution was degraded to $23''$ full width at half power. The r.m.s. noise is $0.5 K km s^{-1}$ in a band $60 km s^{-1}$ wide. The dotted lines mark the border of the area

surveyed. The fields D39, D47 and D84, observed with the interferometer, are indicated by arrows. Throughout this paper we use the M31-fixed coordinate system³ with centre $0^h 40^m 00.3^s$, $41^\circ 00' 03''$ (1950.0), with X along the major axis (PA = 37.7°), positive to the NE, Y along the minor axis, positive to the SE. **b**, $I^*(CO)$ contour levels 2, 4, 6, ... , $16 K km s^{-1}$ superimposed on the Digital Sky Survey (DSS) red image of M31. The linear scale on the right side is for $D = 0.78$ Mpc. The black dot in the bottom left corner represents the telescope $23''$ beam. **c**, $I^*(CO)$ contours smoothed to $24'' \times 36''$, superimposed on the H I map from Brinks and Shane⁴.

that the bulk of the molecular gas in M31's inner disk is too cold to be visible in the CO millimetre lines and that even where CO is bright, the CO brightness distribution might not reflect the true molecular gas distribution.

Extinction of stellar light by dust yields another way of detecting interstellar clouds, which does not depend on temperature or chemistry. Dust grains are known to be closely mixed in space with hydrogen atoms and molecules, and there is a tight correlation¹⁰ in the solar neighbourhood between visual extinction and the total gas column density, $N_{\text{H}} = N(\text{H I}) + 2N(\text{H}_2)$. The dark lanes visible in Fig. 1b reveal many dust-rich clouds located in front of the bright stellar disk. All those with galactocentric radii $R \leq 15$ kpc are detected in CO, including the lanes crossing the dark regions³ D39, D47 and D84, as well as those close to the centre, for example crossing D153. As a rule, CO is brighter where the extinction is larger. The good CO–dust correlation is a very strong indication that CO traces the H_2 column density, as we show below.

The cloud complexes in the disk of M31 are partly molecular, partly atomic, as seen in Fig. 1c, where most clouds stand out in H I and CO. The H I ‘arms’ seem broader than the CO arms, but their crest generally follows the CO arm crest. A notable exception is D39, where the CO emission follows the outer (left) edge of the H I arm and where bright H I emission is detected 30'' to the right of the CO crest, at a place with almost no CO emission. The left border of D39 delineates a 0.5-kpc-wide hole, which hosts a large number of stellar associations (C 116–121). As the bright stars and the H I lie on opposite sides of the CO arm, we conclude that the bulk of the H I there is not a mere photodissociation product of H_2 clouds^{9,11}.

Figure 3a compares point by point the CO and H I velocity-integrated intensities $I(\text{CO})$ and $I(\text{H I})$ inside the arm containing D39 ($R \approx 11$ kpc). Whereas H I is bright in many CO-dim regions with moderate visual extinction, CO is never observed where H I is faint, or the extinction is low. The good correlation between CO and visual extinction implies that a sizable fraction of the gas in the arm is in molecular form.

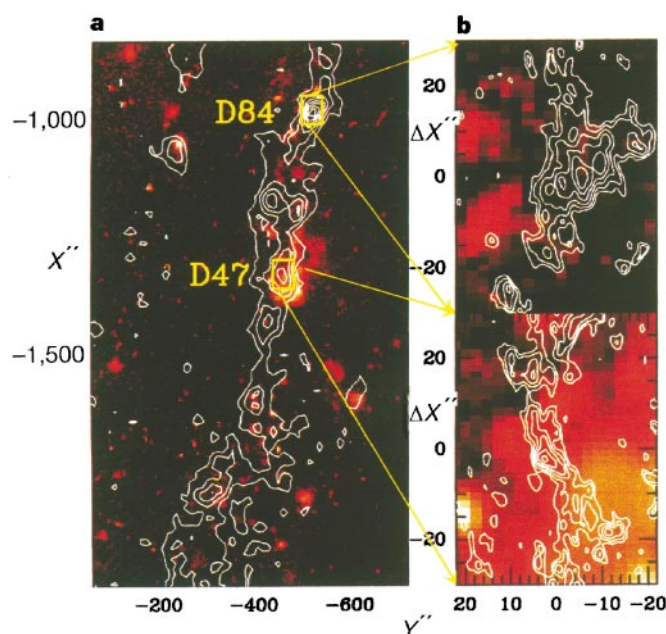


Figure 2 $I(\text{CO})$ contours around D47 and D84, observed with the 30-m telescope (a) and the IRAM interferometer (b), superimposed on an $\text{H}\alpha$ image. H I regions and stars are shown in red. The level step on the interferometer maps is 10 K km s^{-1} . The axes represent offsets along X and Y from the centres of M31, D47 or D84.

To address this issue more quantitatively, we have mapped the 1.2-mm thermal dust emission around D39 (R. Zylka, M. Guélin & N. Neininger, manuscript in preparation). We find a 40 mJy source exactly at the position of the main CO peak. The millimetre dust emission is one of the best tracers of the cold gas^{12,13}, because it is optically thin and scales linearly with the total hydrogen column density N_{H} , the metallicity $Z/Z_{\odot}$, and the average dust temperature, $\langle T_{\text{d}} \rangle$. In M31, at $R = 11$ kpc, $Z/Z_{\odot} \approx 1$ (ref. 14). Averaged over a few hundred parsecs, $\langle T_{\text{d}} \rangle \approx 20$ K in both M31 (ref. 15) and the Milky Way. Adopting a dust absorption cross-section typical of the solar neighbourhood¹², we arrive at $N_{\text{H}} = 10^{22} \text{ cm}^{-2}$ and at a CO-to- H_2 ‘conversion factor’ $\alpha = N(\text{H}_2)/I(\text{CO}) \approx 1.2 \times 10^{20} \text{ cm}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$. In view of the uncertainty ($\sim$ twofold) attached to the dust absorption cross-section, this latter value is marginally smaller than the conversion factor found in the Milky Way¹⁶ ($\alpha = 1.9$ in the above units) and slightly larger than that derived in the nearby spirals NGC 891 (ref. 12) and M51 (ref. 13) ($\alpha = 0.6$ –1). The mass of gas around the peak (area 0.2 arcmin^2) is found to be $1.3 \times 10^6 M_{\odot}$, or $30 M_{\odot} \text{ pc}^{-2}$, and is half atomic, half molecular. The line in Fig. 3a corresponds to $N(\text{H I}) = 2N(\text{H}_2)$, assuming that $\alpha = 1.2$. Only the most obscured lines of sight lie on this line; that is, all others have more atomic than molecular gas.

Maintaining $\alpha = 1.2$, we can estimate N_{H} throughout the 11-kpc arm by using $N_{\text{H}} = 1.8 \times 10^{18} I(\text{H I}) + 2\alpha \times 10^{20} I(\text{CO})$, where we have expressed $N(\text{H I})$ in terms of the H I-line integrated intensity⁵, $I(\text{H I})$. The values of N_{H} are plotted in Fig. 3b against the apparent visual opacity, τ_{R} . The latter is estimated from the optical Digital Sky Survey image, using the expression $I_{\text{obs}}/I_0 = (1 - e^{-\tau})/\tau$, which applies to a thick slab of dust and stars⁴. N_{H} is seen to scale with τ_{R} : the slope is $3 \times 10^{21} \text{ H-cm}^{-2}$, close to that for the clouds in the solar neighbourhood¹⁰. The linear correlation coefficient is $\rho = 0.80$, significantly larger than that obtained by plotting $I(\text{H I})$ or $I(\text{CO})$ alone ($\rho \approx 0.7$), which confirms that both components have comparable masses in the arm crest. Integrated over the entire $R = 11$ -kpc arm, the mass of molecular gas is about half the mass of atomic gas.

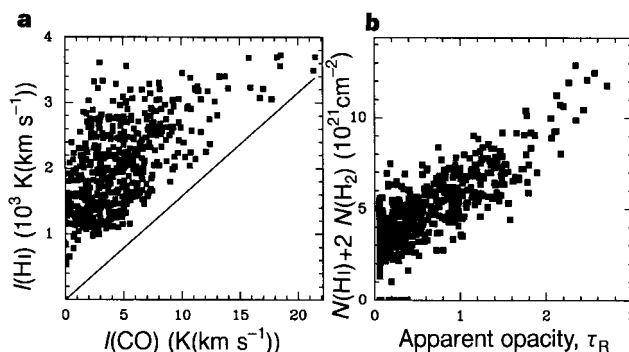


Figure 3 Gas–dust correlation. **a**, Relation between the velocity-integrated intensities corrected for beam efficiency, $I(\text{CO}) = I^*(\text{CO})/0.7$ and $I(\text{H I})$, inside a lane 1.5 kpc wide and 6 kpc long that covers the arm segment associated with D39. The straight line represents equipartition between atomic and molecular gas. The CO data have been smoothed to $24'' \times 36''$. **b**, The total hydrogen density N_{H} as a function of the opacity in the red, τ_{R} . τ_{R} is estimated from the DSS red image of Fig. 1 smoothed to $24'' \times 36''$, using the formula $I_{\text{obs}}/I_0 = (1 - e^{-\tau})/\tau$. I_{obs} is the observed optical intensity (readout from the DSS minus a constant background of 4,700 counts, corresponding to the readouts at the periphery of M31), and I_0 is the expected ‘unobscured’ optical intensity derived in the interarm regions on both sides of the dark lane. The data can be fitted to a straight line with a slope of $3 \times 10^{21} \text{ cm}^{-2}$, which is close to that derived from $\text{L}\alpha$ absorption measurements in the solar neighbourhood¹⁰, once allowance is made for the colour of the DSS image. In fact, $N(\text{H I})$ is $(1$ – $3) \times 10^{21} \text{ cm}^{-2}$ in the interarm region, so that I_0 is slightly underestimated and the straight line does not cross the origin: the true opacity could be somewhat larger than τ_{R} .

Whereas the CO-to-H_I line intensity ratio increases twofold between D39 ($R = 11$ kpc) and the inner arm around D153 ($R = 6$ kpc), the slope of the $I(\text{CO})/\tau_R$ relation remains the same. Yet D153 is not only closer to the centre, but it also shows a much lower star formation activity. Does this mean that Allen and co-workers⁸ are wrong? Their contention that there is a large hidden molecular mass rests on two cloud complexes located at 2 kpc from the centre and might apply only to the innermost region. There, the thick stellar bulge prevents the estimation of τ_R . We note, however, that their mass estimates assume viral equilibrium and a smooth CO brightness distribution, two assumptions that do not seem fulfilled in the largest complex, D478. There, CO profiles⁸ are broad and double-peaked, and look typical of kinematically perturbed clouds near H_I regions (such as D47 in Fig. 2b). They also show evidence of CO brightness fluctuations on scales of 40 pc. Thus, even in these innermost regions, the evidence for hidden molecular gas is meagre.

Because CO is a good tracer of H₂, Fig. 1a yields the molecular cloud distribution in M31 with an unprecedented resolution. The brightest clouds are grouped in narrow filamentary arms, which stand out against a rich background of fainter clouds. The connection of the arm segments into a spiral pattern is not straightforward. One solution is illustrated by the three ellipses of Fig. 1a. It assumes a major axis position angle $\text{PA} = 42^\circ$, the value that fits the southern velocity field best, and an inclination on the plane of the sky $i = 78^\circ$. In this picture, all the bright CO clouds lie on three circular arms of radii $R = 7, 11$ and 18 kpc and on the border of the stellar super-association NGC 206. Another arm pattern¹⁷ assumes that $\text{PA} = 38^\circ$ and two logarithmic spirals with a pitch angle of 7° .

The CO arms are very narrow, as illustrated in Fig. 2, which shows the cloud complexes associated with D47 and D84, observed with resolution of $2''$ (9 pc). The bright blobs of Fig. 1 are now resolved into thin, clumpy filaments quite similar in brightness and size to Orion A and Orion B¹⁸; the peak CO brightness temperature, averaged over the ~ 9 pc beam, is 7 K. We note that, although D84 is the strongest CO source in the southwest half of M31, Orion is not the brightest source by far in the Milky Way, but only the brightest in the vicinity of the Sun. Obviously, M31 lacks a bright inner structure like the Galactic 'Molecular Ring', but it shows relatively strong molecular emission up to $R = 18$ kpc, farther than the Milky Way¹⁹.

Does CO emission trace the molecular gas in other galaxies as well as in M31? A correlation between CO arms and dark dust lanes is also observed^{11,20} in the Sc-type spiral M51, but on a coarser scale. M51, 10 Mpc away, is the archetype of spirals with a high star-formation rate. Poorer resolution and a less uniform stellar background make it difficult to repeat our analysis, but again, all conspicuous dark lanes of M51 seem to have bright CO counterparts. The millimetre dust emission has been mapped in M51 (ref. 13) and in half a dozen of Sb–Sd spirals, including NGC 891 (ref. 12), which is often considered a twin of the Milky Way. Except in the outermost parts of the disks, there is a tight correlation between this emission and CO. Although the dust emission depends on temperature, this tight correlation strongly suggests that CO is indeed a fair tracer for the molecular clouds in most spiral galaxies. □

Received 9 January; accepted 8 July 1998.

- Maloney, P. & Black, J. H. Molecular gas abundances in galaxies. *Astrophys. J.* **325**, 389–401 (1988).
- Stanek, K. Z. & Garnavich, P. M. Distance to M31 with the *HST* and *Hipparcos* Red Clump Stars. *Astrophys. J. Lett.* **503**, L131–L134 (1998).
- Hodge, P. *The Andromeda Galaxy* Ch. 10–12 (Astrophysics and Space Science Library, Vol. 176, Kluwer, Dordrecht, 1992).
- Walterbos, R. A. M. & Kennicutt, R. C. Jr An optical study of stars and dust in the Andromeda galaxy. *Astron. Astrophys.* **198**, 61–86 (1988).
- Brinks, E. & Shane, W. W. A high resolution H_I line survey of M 31. *Astron. Astrophys. Suppl.* **55**, 179–251 (1984).
- Dame, T. M. *et al.* A complete CO survey of M 31. *Astrophys. J.* **418**, 730–742 (1993).
- Loiseau, L. *et al.* Molecular spiral arms in M 31. *Astrophys. J.* **469**, L101–L104 (1996).
- Loiseau, L. & Allen, R. J. Cold massive molecular clouds in the inner disk of M 31. *Astrophys. J.* **499**, 227–233 (1998).
- Tilanus, R. P. J. & Allen, R. J. Spiral structure of M 83. *Astron. Astrophys.* **274**, 707–729 (1993).
- Böhlin, R. C., Savage, B. D. & Drake, J. F. A survey of interstellar H_I from $\text{L}\alpha$ absorption measurements. *Astrophys. J.* **224**, 132–142 (1978).

- Rand, R. J., Kulkarni, S. R. & Rice, W. Star formation and the distribution of H_I and infrared emission in M 51. *Astrophys. J.* **390**, 66–78 (1992).
- Guélin, M. *et al.* 1.3 mm emission in the disk of NGC 891: evidence of cold dust. *Astron. Astrophys.* **279**, L37–L40 (1993).
- Guélin, M. Cold dust emission from the spiral arms of M 51. *Astron. Astrophys.* **298**, L29–L32 (1995).
- Blair, W. P., Kirshner, R. P. & Chevalier, R. A. Abundance gradients in M 31. *Astrophys. J.* **254**, 50–69 (1982).
- Walterbos, R. A. M. & Schwering, P. B. W. Infrared emission from interstellar dust in the Andromeda Galaxy. *Astron. Astrophys.* **180**, 27–49 (1987).
- Strong, A. W. & Mattox, J. R. Gradient model analysis of EGRET diffuse Galactic γ -ray emission. *Astron. Astrophys.* **308**, L21–L24 (1996).
- Braun, R. Distribution and kinematics of neutral gas in M 31. *Astrophys. J.* **372**, 54–66 (1991).
- Maddalena, R. J. *et al.* The large system of molecular clouds in Orion and Monoceros. *Astrophys. J.* **303**, 375–391 (1986).
- Heyer, M. *et al.* the FCRAO CO survey of the Outer Galaxy. *Astrophys. J. Suppl.* **115**, 241–258 (1998).
- García-Burillo, S., Guélin, M. & Cernicharo, J. CO in M 51. I. Molecular spiral structure. *Astron. Astrophys.* **274**, 123–147 (1993).
- Emerson, D. T. & Gräve, R. The reduction of scanning noise in raster scanned data. *Astron. Astrophys.* **190**, 353–358 (1988).
- Hoernes, P. Thesis, Univ. Bonn (1997).

Acknowledgements. We thank the IRAM Granada staff, particularly A. Sievers and W. Brunswig, and P. Hoenes from MPIfR for their important role in developing the 'on-the-fly' observation and data-reduction procedures. We also thank E. Brinks and N. Devereux for making available their H_I and H₂ data.

Correspondence and requests for materials should be addressed to N.N. (e-mail: nneini@astro.uni-bonn.de) and M.G. (e-mail: guelin@iram.fr).

Microwave spectroscopy of a quantum-dot molecule

T. H. Oosterkamp*, T. Fujisawa**†, W. G. van der Wiel*, K. Ishibashi*‡, R. V. Hijman*, S. Tarucha† & L. P. Kouwenhoven*

* Department of Applied Physics and DIMES, Delft University of Technology, PO Box 5046, 2600 GA Delft, The Netherlands

† NTT Basic Research Laboratories, 3-1, Morinosoto Wakamiya, Atsugi-shi, Kanagawa 243-0198, Japan

‡ Institute of Physical and Chemical Research (RIKEN), 2-1, Hirosawa, Wako, Saitama 351-01, Japan

Quantum dots are small conductive regions in a semiconductor, containing a variable number of electrons (from one to a thousand) that occupy well-defined, discrete quantum states—for which reason they are often referred to as artificial atoms¹. Connecting them to current and voltage contacts allows the discrete energy spectra to be probed by charge-transport measurements. Two quantum dots can be connected to form an 'artificial molecule'. Depending on the strength of the inter-dot coupling (which supports quantum-mechanical tunnelling of electrons between the dots), the two dots can form 'ionic' (refs 2–6) or 'covalent' bonds. In the former case, the electrons are localized on individual dots, while in the latter, the electrons are delocalized over both dots. The covalent binding leads to bonding and antibonding states, whose energy difference is proportional to the degree of tunnelling. Here we report a transition from ionic bonding to covalent bonding in a quantum-dot 'artificial molecule' that is probed by microwave excitations^{5–8}. Our results demonstrate controllable quantum coherence in single-electron devices, an essential requirement for practical applications of quantum-dot circuitry.

When particles are allowed to tunnel back and forth between two quantum systems, the energy states of the individual systems mix and form new states that extend over both systems. The extended states are referred to as the bonding or symmetric state, and the antibonding or antisymmetric state. In solid-state systems, the energy splitting between bonding and antibonding states have been observed in quantum-well structures^{9,10}, superconducting tunnelling devices^{11,12}, and exciton systems¹³.

Quantum dots are uniquely engineered solid-state systems in the

sense that they have discrete states and the electrons on the dots are strongly interacting. The question whether different dots can be coupled together in a quantum-mechanically coherent way is non-trivial. The reason is that quantum dots composing single-electron devices are embedded in an environment with many electronic degrees of freedom. The electron that occupies the covalent state of a double-dot system has a Coulomb interaction with all the other electrons confined on the dots and also with the electrons in the current and voltage leads. These interactions can lead to dephasing of the quantum-mechanical wavefunction resulting in a breakdown of the covalent state. For realistic devices there is yet no theory that can calculate reliable dephasing rates. Nevertheless, if elements like quantum dots are ever to be integrated in little quantum circuits^{14–16}, it is necessary that dots can be coupled coherently.

We have used microwave spectroscopy (0–50 GHz) to measure the energy differences between states in the two dots of the device² shown in Fig. 1a. We show that these energy differences, including

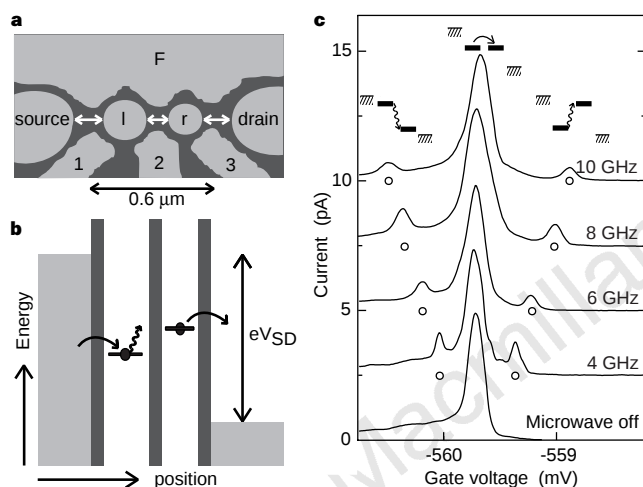


Figure 1 Photon resonances in a double-dot sample. **a**, Photograph of the double quantum dot sample. The source and drain regions as well as the left and right dots are indicated schematically. The tunnel barriers are depicted as arrows. The metallic gates (1, 2, 3 and F) are fabricated on top of a GaAs/AlGaAs heterostructure with a two-dimensional electron gas (2DEG) 100 nm below the surface. At 4.2 K the 2DEG mobility is $2.3 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and the electron density is $1.9 \times 10^{15} \text{ m}^{-2}$. Applying negative voltages to all the gates depletes the electron gas underneath them and forms two dots with estimate sizes of $170 \text{ nm} \times 170 \text{ nm}$ and $130 \text{ nm} \times 130 \text{ nm}$. We measure the d.c. photocurrent in response to a microwave signal (0–50 GHz) that is capacitively coupled to gate 2. The tunnel coupling between the two dots and to the reservoirs can be controlled with the voltages on gate 1, 2 and 3. The dots left and right contain about 60 and 35 electrons, respectively. The sample is cooled in a dilution refrigerator, yielding an electron temperature in the source and drain contacts of $\sim 100 \text{ mK}$. **b**, Diagram of the electron energies in the dot for the case that an electron needs to absorb a photon in order to contribute to the current. Shaded areas represent the electron states in the leads that are continuously filled up to the Fermi levels. A voltage V_{SD} applied between the source and drain contacts shifts one Fermi level relative to the other. The discrete energy states in the two dots can be adjusted independently by changing the gate voltages. **c**, The upper diagrams illustrate three situations of the energy state in the left dot relative to the state in the right dot. The hatched lines denote the Fermi levels in the leads. The bottom curve shows the current as a function of the voltage on gate 1 for $V_{\text{SD}} = 500 \text{ } \mu\text{V}$ without applying microwaves. A single resonance occurs when two states line up. Other curves, which have been offset for clarity, show the current when microwaves with frequency f from 4 to 10 GHz are applied. Now, two additional satellite resonances occur when the two states are exactly a photon energy apart. The corresponding photon-assisted tunnelling processes are illustrated in the upper diagrams.

the bonding–antibonding splitting, is controlled by gate voltages which tune the tunnel coupling between the dots. We first discuss the weak-coupling regime.

Electrons are strongly localized on the individual dots when tunnelling between the two dots is weak. Electron transport is then governed by single-electron charging effects⁸. The charging energies can be tuned away by means of the gate voltages. It is then energetically allowed for an electron to tunnel between dots when a discrete state in the left dot is aligned with a discrete state in the right dot. External voltages also control the alignment of the discrete states. A current can flow when electrons can tunnel, while conserving energy, from the left lead, through the left and right dots, to the right lead. We note that energy is also conserved when a photon of energy hf (where h is Planck's constant), which matches the energy difference between the states of the two dots, is absorbed from the microwave field of frequency f (Fig. 1b).

The resonance in the lowest trace in Fig. 1c is due to an alignment of discrete states. The other traces are measured while applying a microwave signal. The satellite resonances are due to photon-assisted tunnelling processes which involve the emission (left satellite) or absorption (right satellite) of a microwave photon.

Stoof and Nazarov¹⁷ give a detailed description of photon-assisted tunnelling in a double quantum dot. The basic idea is that electrons can absorb fixed quanta of energy hf from a classical oscillating field. An a.c. voltage drop $V = V_{\text{ac}} \cos(2\pi ft)$ across a tunnel barrier modifies the tunnel rate through the barrier as¹⁸:

$$\tilde{\Gamma}(E) = \sum_{n=-\infty}^{+\infty} J_n^2(\alpha) \Gamma(E + nhf) \quad (1)$$

Here $\tilde{\Gamma}(E)$ and $\Gamma(E)$ are the tunnel rates at energy E with and without an a.c. voltage, respectively; $J_n^2(\alpha)$ is the square of the n th

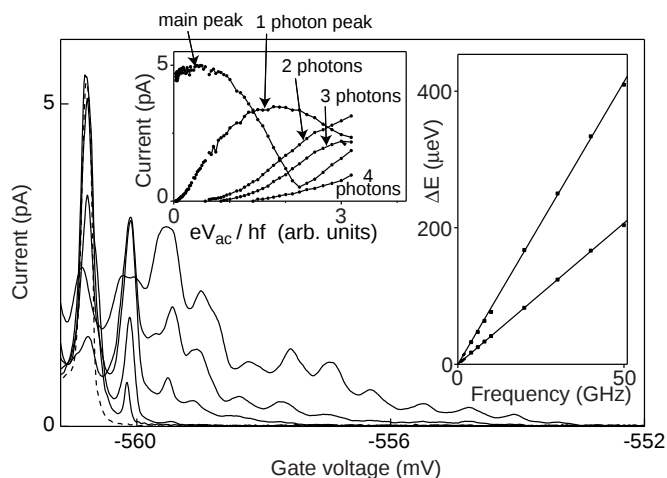


Figure 2 Current versus gate voltage of a weakly coupled double-dot. The dashed curve is without microwaves and contains only the main resonance. The solid curves are taken at 8 GHz for increasing microwave powers resulting in an increasing number of satellite peaks. At the right side of the main peak these correspond to photon absorption. The source–drain voltage $V_{\text{SD}} = 700 \text{ } \mu\text{V}$ and the photon energy $hf = 32 \text{ } \mu\text{eV}$ at 8 GHz. At the highest power we observe 11 satellite peaks, demonstrating multiple photon absorption. Left inset, height of the first four satellite peaks as a function of the microwave amplitude. The observed height dependence agrees with the expected Bessel function behaviour. Right inset, distance between main resonance and first two satellites as a function of the applied frequency from 1 to 50 GHz. The distance is transferred to energy through $\Delta E = \kappa \Delta V_g$ where κ is the appropriate capacitance ratio for our device that converts gate voltage V_g to energy⁸. The agreement between data points and the two solid lines, which have slopes of h and $2h$, demonstrates that we observe the expected linear frequency dependence of the one- and two-photon processes.

order Bessel function evaluated at $\alpha = (eV_{ac})/(hf)$, which describes the probability that an electron absorbs or emits n photons of energy hf . e is the electron charge.

Figure 2 shows the current for several microwave powers. The dashed curve shows the main resonance measured at zero power. As the power is increased, satellite peaks appear corresponding to the absorption of multiple photons which are observed up to $n = 11$. At these high powers, the microwaves strongly perturb tunnelling. This is reflected by the nonlinear dependence of the peak heights on power (left inset of Fig. 2), which is in agreement with the expected Bessel-function behaviour.

The right inset to Fig. 2 shows that the separation of the satellite peaks from the main peak depends linearly on frequency between 1 and 50 GHz. As we discuss below, this linearity implies that the tunnel coupling is negligible. The electrons are thus localized on the individual dots and they have an ionic bonding. The line proportional to $2hf$ is taken from data at higher microwave powers where electrons absorb or emit two photons during tunnelling.

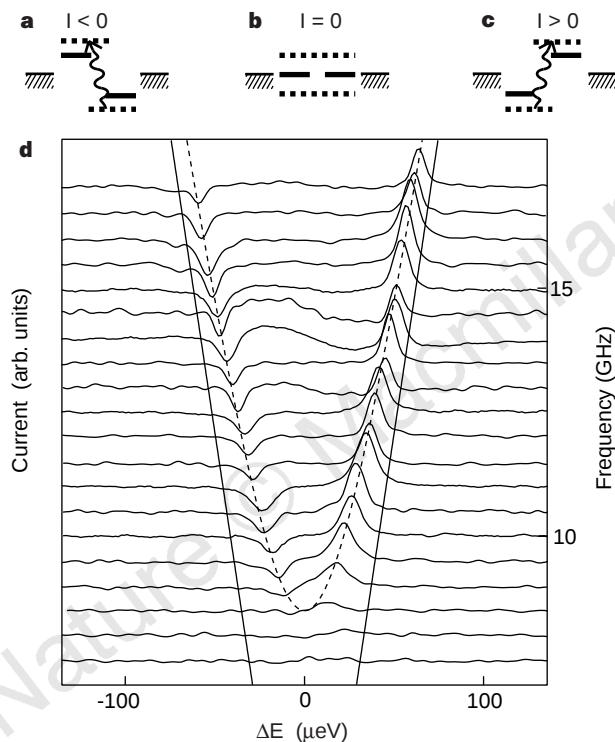


Figure 3 Measured pumped current through the strongly coupled double-dot. **a–c**, Energy diagrams. Solid lines depict the energy states E_{left} and E_{right} in the two dots for the case that the coupling is weak and that their energy difference is simply $\Delta E = E_{\text{left}} - E_{\text{right}}$. When the dots are strongly coupled, the states delocalize over both dots thereby forming a bonding and an antibonding state. These are indicated by two dotted lines. Their energy difference is $\Delta E^* = [\Delta E^2 + (2T)^2]^{1/2}$. Electrons are transferred from the bonding to the antibonding state when $\Delta E^* = hf$. In **a**, $E_{\text{left}} > E_{\text{right}}$ which results in electron pumping from right to left corresponding to a negative current. In **b**, the whole system is symmetric ($E_{\text{left}} = E_{\text{right}}$) and consequently the net electron flow must be zero. In **c**, $E_{\text{left}} < E_{\text{right}}$ which gives rise to pumping from left to right and a positive current. **d**, Measured pumped current through the strongly coupled double-dot. Gates 1 and 3 are swept simultaneously in such a way that we vary the energy difference ΔE . The different traces are taken at different microwave frequencies, and are offset such that the right vertical axis gives the frequency. The main resonance is absent as we have set $V_{\text{SD}} = 0$. The satellite peaks typically have an amplitude of 0.5 pA. For weakly coupled dots the satellite peaks are expected to move linearly with frequency, thereby following the straight solid lines. In contrast, we observe that the satellite peaks follow the fitted dotted hyperbola $hf = [\Delta E^2 + (2T)^2]^{1/2}$ using T as a fitting parameter.

In contrast to the case of weakly coupled dots, covalent bonding occurs when two discrete states that are spatially separated become strongly coupled. Electrons then tunnel quickly back and forth between the dots. In a quantum-mechanical description this results in a bonding and antibonding state which are respectively lower and higher in energy than the original states. Our strong-coupling measurements were made on a second type of double-dot sample (see inset to Fig. 4). To single out the current only due to microwaves we operate the device as an electron pump driven by photons in a way described theoretically by Stafford and Wingreen¹⁹ and by Brune *et al.*²⁰ (see the diagrams of Fig. 3a–c). By sweeping the gate voltages we vary $\Delta E = E_{\text{left}} - E_{\text{right}}$, where E_{left} and E_{right} are the energies of the uncoupled states in the left and right dot. The bonding and antibonding states, that are a superposition of the wavefunctions corresponding to an electron in the left or in the right dot, have an energy splitting of $\Delta E^* = E_{\text{antibond}} - E_{\text{bond}} = [(\Delta E)^2 + (2T)^2]^{1/2}$, where T is the tunnel coupling between the two dots. When the sample is irradiated, a photocurrent may result as illustrated in Fig. 3a–c. A non-zero current indicates that an electron was excited from the bonding state to the antibonding state, thereby fulfilling the condition $hf = \Delta E^*$, or conversely:

$$\Delta E = \sqrt{(hf)^2 - (2T)^2} \quad (2)$$

Figure 3 shows measured current traces as a function of the uncoupled energy splitting ΔE , where from top to bottom the

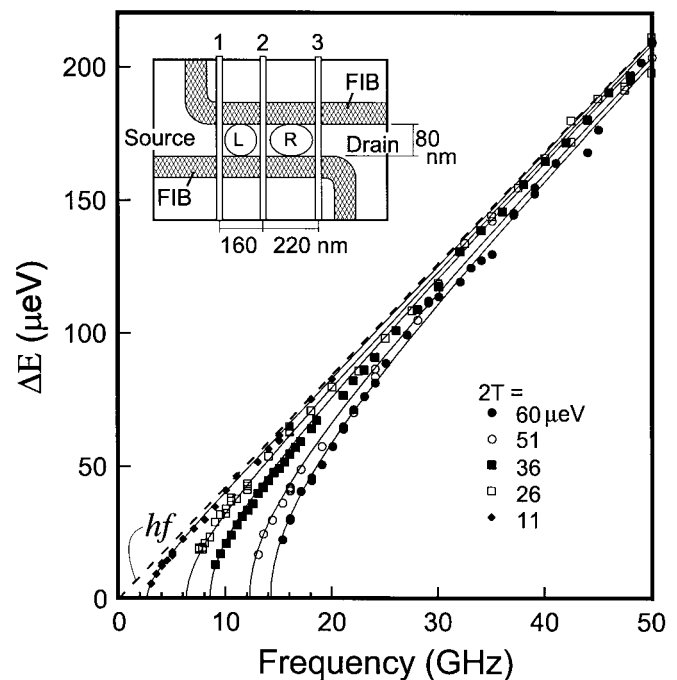


Figure 4 Half the spacing in gate voltage between the positive and negative satellite peaks as a function of frequency. Gate voltage spacing has been transferred to energy difference ΔE (see also Fig. 2 legend). Different curves correspond to different coupling constants T . Solid lines are theoretical fits to $\Delta E = [(\Delta E)^2 + (2T)^2]^{1/2}$. The resulting values for $2T$ are given in the figure. In the limit of weak coupling, this reduces to $\Delta E = hf$ which is indicated by the dashed line. The coupling is varied by applying different voltages to the centre gate (2) or by changing the magnetic field (filled diamonds, $B = 3.3$ T; filled squares, $B = 2.2$ T; other curves, $B = 0$). The upper left inset shows a diagram of the sample⁵. A narrow channel is defined by locally depleting the 2DEG using focused ion beam implantation (FIB). Two dots are then formed by applying negative voltages to the three gates (1, 2, 3) that cross the channel. Microwaves are capacitively coupled to gate 2.

applied microwave frequency is decreased from 17 to 7.5 GHz in 0.5 GHz steps. The distance between the pumping peaks, which is proportional to $2\Delta E$, decreases as the frequency is lowered. However, the peak distance decreases faster than linearly with frequency; the peaks follow the hyperbola rather than the straight lines. The distance goes to zero when the frequency approaches the minimum energy gap between bonding and antibonding states, $hf = 2T$. For frequencies smaller than the coupling, $hf < 2T$, the photon energy is too small to induce a transition from the bonding to the antibonding state.

The coupling between the dots can be decreased by changing the gate voltage on the centre gate to more negative values, or by applying a magnetic field perpendicular to the sample. In Fig. 4 we have plotted the frequency dependence of the energy spacing ΔE at which the pumping current is at a maximum. Different plotting symbols correspond to different centre gate voltage settings and magnetic fields. The solid lines are fits of equation (2) to the measured data. It follows that the coupling $2T$ has been tuned from 11 to 60 μeV . The good agreement with equation (2) and the clear nonlinear frequency dependence demonstrates the control over the formation of covalent bonding between the two dots.

Quantum dots have been suggested as possible candidates for building a quantum computer^{14–16}. We have shown that it is indeed possible to coherently couple dots, and that one can induce transitions between the extended states. The next crucial step towards quantum logic gates is to show that the coherence of the superposition is preserved on timescales much longer than the time needed for manipulating the electron wavefunctions. A lower bound for the dephasing time is 1 ns, which we deduce from our narrowest peaks and from the smallest energy gaps between the bonding and antibonding states that we have resolved. We intend to perform measurements of the decoherence time in which the states are manipulated by applying the microwaves in short pulses. $\square$

Received 15 April; accepted 15 July 1998.

1. Ashoori, R. Electrons in artificial atoms. *Nature* **379**, 413–419 (1996).
2. van der Vaart, N. C. *et al.* Resonant tunneling through two discrete energy states. *Phys. Rev. Lett.* **74**, 4702–4705 (1995).
3. Livemore, C. *et al.* The Coulomb blockade in coupled quantum dots. *Science* **274**, 1332–1335 (1996).
4. Blick, R. *et al.* Single electron tunneling through a double quantum dot: The artificial molecule. *Phys. Rev. B* **53**, 7899–7902 (1996).
5. Fujisawa, T. & Tarucha, S. Photon assisted tunneling in single and coupled quantum dot systems. *Superlatt. Microstruct.* **21**, 247–254 (1997).
6. Fujisawa, T. & Tarucha, S. Multiple photon assisted tunneling between two coupled quantum dots. *Jpn J. Appl. Phys.* **36**, 4000–4003 (1997).
7. Oosterkamp, T. H. *et al.* Photon sidebands of the ground state and first excited state of a quantum dot. *Phys. Rev. Lett.* **78**, 1536–1539 (1997).
8. Kouwenhoven, L. P. *et al.* in *Mesoscopic Electron Transport* (eds Sohn, L. *et al.*) 105–214 (NATO ASI Ser. E Vol. 345, 1997); also (<http://vortex.tn.tudelft.nl/~leek/papers>).
9. Leo, K. *et al.* Coherent oscillations of a wave packet in a semiconductor double-quantum well structure. *Phys. Rev. Lett.* **66**, 201–204 (1991).
10. Mendez, E. E. & Bastard, G. Wannier-Stark ladders and Bloch oscillations in superlattices. *Phys. Today* **46**, 34–42 (1993).
11. Flees, D. J. *et al.* Interband transitions and band gap measurements in Block transistors. *Phys. Rev. Lett.* **78**, 4817–4820 (1997).
12. Nakamura, Y. *et al.* Spectroscopy of energy-level splitting between two macroscopic quantum states of charge coherently superposed by Josephson coupling. *Phys. Rev. Lett.* **79**, 2328–2331 (1997).
13. Schedelbeck, G. *et al.* Coupled quantum dots fabricated by cleaved edge overgrowth: From artificial atoms to molecules. *Science* **278**, 1792–1795 (1997).
14. Landauer, R. Minimal energy requirements in communication. *Science* **272**, 1914–1918 (1996).
15. Barenco, A. *et al.* Conditional quantum dynamics and logic gates. *Phys. Rev. Lett.* **74**, 4083–4086 (1995).
16. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
17. Stoof, T. H. & Nazarov, Yu. V. Time-dependent resonant tunneling via two discrete states. *Phys. Rev. B* **53**, 1050–1053 (1996).
18. Tien, P. K. & Gordon, J. R. Multiphoton process observed in the interaction of microwave fields with the tunneling between superconductor films. *Phys. Rev.* **129**, 647–651 (1963).
19. Stafford, C. A. & Wingreen, N. S. Resonant photon-assisted tunneling through a double quantum dot: An electron pump from spatial Rabi oscillations. *Phys. Rev. Lett.* **76**, 1916–1919 (1996).
20. Brune, Ph. *et al.* Photon-assisted transport through a double quantum dot with a time-dependent interdot barrier. *Physica E* **1**, 216–218 (1997).

Acknowledgements. We thank R. Aguado, S. M. Cronenwett, S. F. Godijn, P. Hadley, C. J. P. M. Harmans, K. K. Likharev, Y. E. Mooij, Yu. V. Nazarov, R. M. Schouten, T. H. Stoof and N. C. van der Vaart for experimental help and discussions. This work was supported by the Dutch Organization for Research on Matter (FOM) and by the EU via the TMR network. L.P.K. was supported by the Dutch Royal Academy of Arts and Sciences (KNAW).

Correspondence and requests for materials should be addressed to T.H.O. (e-mail: tjerk@qt.tn.tudelft.nl).

Localized vibrational modes in metallic solids

V. Keppens*, D. Mandrus*, B. C. Sales*, B. C. Chakoumakos*, P. Dai*, R. Coldea*, M. B. Maple†, D. A. Gajewski†, E. J. Freeman† & S. Bennington‡

* Solid State Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

† Department of Physics and Institute for Pure and Applied Physical Sciences, University of California San Diego, La Jolla, California 92093, USA

‡ ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, UK

Filled skutterudite antimonides^{1,2} are cubic compounds with the formula $\text{RM}_4\text{Sb}_{12}$, where R is a rare-earth element (such as La or Ce), and M is a transition metal (for example, Fe or Co). The rare-earth ion is weakly bound in an oversized atomic cage formed by the other atoms. Its presence has been shown to cause a dramatic reduction in the lattice component of the thermal conductivity, while having little effect on the electronic properties^{3–5} of the compound. This combination of properties makes filled skutterudites of interest as thermoelectric materials. It has been suggested⁴ that localized, incoherent vibrations of the rare-earth ion are responsible for the reduction in thermal conductivity, but no direct evidence for these local vibrational modes exists. Here we report the observation of local modes in La-filled skutterudites, using heat capacity, elastic constant and inelastic neutron scattering measurements. The La atoms show unusual thermodynamic behaviour, characterized by the presence of two low-energy localized modes. Our results suggest that consideration of local modes will play an important role in the design of the next generation of thermoelectric materials.

Localized vibrational modes are uncommon in solids because of the strong interactions that exist between the constituent atoms. When present, local modes are usually associated with weakly bound guest atoms that reside in the voids of an open-structured non-metallic host. In metallic solids, which tend to crystallize in close-packed structures, local modes are exceedingly rare. These low-energy vibrational modes, which are not present in the unfilled parent compound CoSb_3 , give unambiguous evidence for the ‘rattling’ behaviour of the rare-earth atom in the

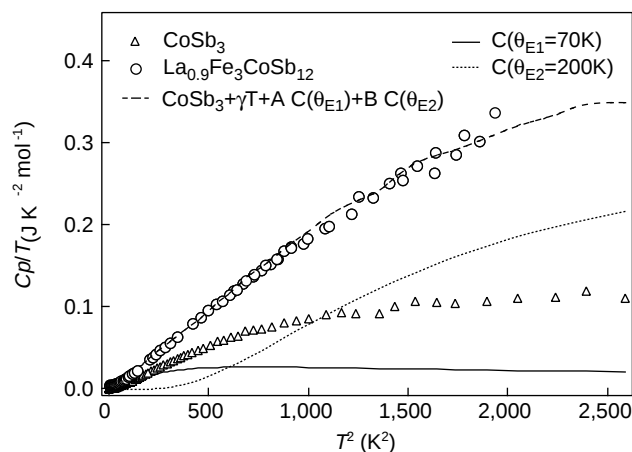


Figure 1 Specific heat divided by temperature versus temperature squared for $\text{La}_{0.9}\text{Fe}_3\text{CoSb}_{12}$ and CoSb_3 . The dashed line through the $\text{La}_{0.9}\text{Fe}_3\text{CoSb}_{12}$ data is based on a calculation in which the contribution from two Einstein oscillators with level spacings of 70 and 200 K are added to the CoSb_3 data (see text for details).

structure. Because filled skutterudites display a wide variety of ground states, including metallic³, semimetallic⁶, semiconducting⁴, superconducting⁷, Kondo insulating⁸ and ferromagnetic (both metallic⁹ and insulating⁸), filled skutterudites offer a means to study the effects of localized atomic vibrations on a wide variety of physical phenomena.

Polycrystalline specimens were prepared by melting stoichiometric quantities of high-purity elements in carbon-coated, sealed and evacuated silica tubes. The details of the synthesis are described in ref. 4. Specific heat measurements were performed from 2 to 45 K for CoSb₃ and La_{0.9}Fe₃CoSb₁₂. The results are plotted in Fig. 1, which illustrates the significant difference in the low-temperature behaviour of the filled and unfilled materials. Although the specific heat of CoSb₃ deviates somewhat from simple Debye behaviour owing to the presence of some low-energy optical phonons involving collective motions of 4-membered Sb rings, a model proposed by Feldman and Singh¹⁰ is able to quantitatively account for the data (J. Feldman, personal communication). Using the CoSb₃ data as a background, we find that the addition of two quantized harmonic oscillators (Einstein oscillators) is required in order to model the specific heat of La_{0.9}Fe₃CoSb₁₂. As illustrated in Fig. 1, a model calculation including the contribution of two Einstein oscillators leads to an adequate description of the experimental results. The dashed line through the La_{0.9}Fe₃CoSb₁₂ data in Fig. 1 represents a fit to the equation:

$$C_p = C_p(\text{CoSb}_3) + \gamma T + AC_{E1} + BC_{E2}$$

with $C_p(\text{CoSb}_3)$ the molar specific heat for CoSb₃, $\gamma = 0.0037 \text{ mol}^{-1} \text{ K}^{-2}$, $A = 1.2 \text{ mol}^{-1} \text{ K}^{-1}$, $B = 35.0 \text{ mol}^{-1} \text{ K}^{-1}$ and T is temperature. Here C_{E1} and C_{E2} represent contributions from Einstein oscillators, $C_E = (\theta/T)^2 e^{(\theta/T)} / (e^{(\theta/T)} - 1)^2$, with Einstein temperatures $\theta_{E1} = 70 \text{ K}$ and $\theta_{E2} = 200 \text{ K}$. The fact that two Einstein oscillators are required to model the data indicates that two distinct local modes may exist in the La-filled skutterudite. It is tempting to ascribe the lower-energy oscillator to the rattling of the La ions

because the magnitude of the 70 K oscillator is about what we expect for a single ion oscillating in a harmonic potential. At high temperatures, we expect each mole of oscillators to contribute $3R = 24.95 \text{ J K}^{-1}$ to the heat capacity. As only $\sim 5.3\%$ of the atoms rattle, we anticipate a contribution of $\sim 1.33 \text{ mol}^{-1} \text{ K}^{-1}$. This is very close to the fitted value of $1.21 \text{ mol}^{-1} \text{ K}^{-1}$. We note that the results of the analysis are quite similar if one models the specific heat of La_{0.9}Fe₃CoSb₁₂ using a T^3 Debye term instead of the measured specific heat of CoSb₃. In that case two Einstein oscillators are again required, this time with $\theta_{E1} = 70 \text{ K}$ and $\theta_{E2} = 157 \text{ K}$.

To investigate further the unusual thermodynamics of the La-filled skutterudite, we measured the elastic moduli of the filled and unfilled materials as a function of temperature. High-quality samples of La_{0.75}Fe₃CoSb₁₂, with a density that is 98% of the theoretical value, and CoSb₃ (95% dense) were cut into $2 \times 2.5 \times 3 \text{ mm}$ rectangular parallelepipeds and used for the resonant ultrasound spectroscopy (RUS) measurements. RUS is an unusual ultrasonic technique, developed by Migliori *et al.*¹¹ for determining the complete set of elastic moduli of a solid by measuring the free-body resonances of the sample. This method is unique in that all moduli can be determined simultaneously, avoiding remounts of transducers and multiple temperature sweeps. An isotropic polycrystalline solid has two elastic moduli, a shear modulus c_{44} governing transverse waves, and a compressional modulus c_{11} governing longitudinal waves. Measurements have been performed as a function of temperature (5–300 K) for both filled and unfilled skutterudite specimens. The upper panel of Fig. 2 shows the shear modulus c_{44} for La_{0.75}Fe₃CoSb₁₂, and in the inset the modulus for the unfilled skutterudite CoSb₃. The solid line through the CoSb₃ data is a model calculation, using the so-called Varshni function¹²:

$$c_{ij}(T) = c_{ij}^0 - s/(e^{T/\theta} - 1)$$

with T the temperature, c_{ij}^0 the elastic constant at 0 K, and s and θ fitting parameters. This function was shown by Varshni to describe the temperature dependence of the elastic constants of many simple substances, and characterizes to some extent 'normal' elastic behaviour. The elastic response of CoSb₃ (Fig. 2) can be well-described by the Varshni model. However, the shear modulus of the filled skutterudite La_{0.75}Fe₃CoSb₁₂ behaves anomalously at low temperatures,

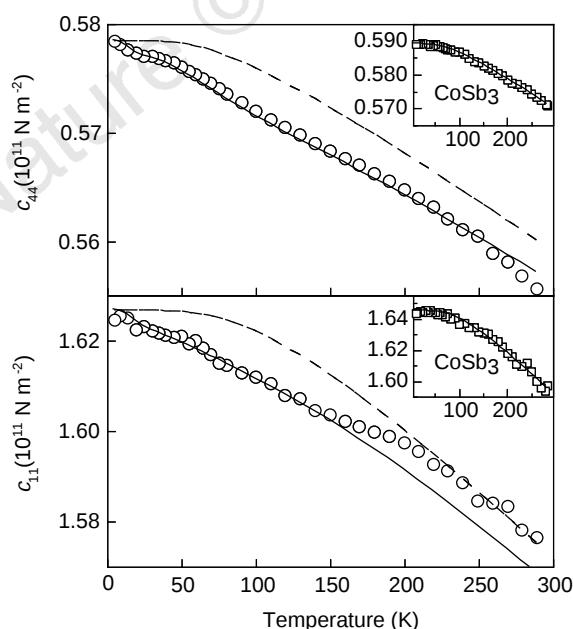


Figure 2 Elastic moduli c_{11} and c_{44} for La_{0.75}Fe₃CoSb₁₂ as a function of temperature. The solid lines through the data represent model calculations using two two-level systems; one with a level spacing of 50 K and the other with spacing of 200 K. The dashed lines are estimated background elastic moduli. The "normal" elastic behaviour of CoSb₃ is shown in the insets; the quantities plotted on the axes are the same as those in the main figures.

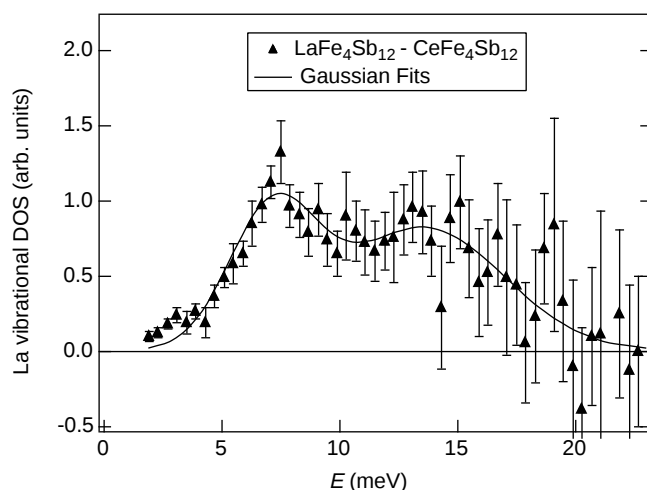


Figure 3 Difference in the inelastic neutron scattering data between LaFe₄Sb₁₂ and CeFe₄Sb₁₂ versus energy loss. The incident neutron energy was 30 meV and the energy resolution was $\sim 2 \text{ meV}$. CeFe₄Sb₁₂ was used as a reference because the neutron scattering cross-section of Ce is much smaller than that of La. The difference spectra therefore reflect the vibrational density of states (DOS) associated with the La atoms. The peaks at 7 and 15 meV correspond to temperatures of 80 and 175 K.

displaying a much stronger temperature dependence, which is suggestive of unusual low-energy vibrational modes in addition to the normal acoustic phonons.

A first attempt to model this unusual behaviour, taking into account the presence of harmonic Einstein-oscillators, failed to describe the 'dip' in the data at low temperatures. The elastic data can, however, be described by considering the elastic response of a two-level system (TLS) with level-spacing Δ . We can calculate the TLS-contribution to the elastic response, $c = \partial^2 F / \partial \epsilon^2$, with $F = -N_A k_B T \ln(1 + e^{-\Delta/T})$ the Helmholtz free energy of a TLS, and assuming that the vibrational level spacing, Δ , depends linearly on the strain, ϵ , that is, $\Delta = \Delta_0 + A\epsilon$, with A a coupling constant. As illustrated in Fig. 2, both the c_{11} and c_{44} modulus can be modelled reasonably well by adding two TLSs, with level spacings of 50 and 200 K. The dashed line in Fig. 2 represents the 'background' contribution, which is estimated from the Varshni-fit for the unfilled skutterudite CoSb₃.

As the TLS approach quite effectively models the RUS data, whereas the Einstein modes fail, we investigated whether the specific heat, modelled above with two Einstein oscillators, could be described using two-level systems. We found that the data can indeed be modelled using two TLSs, with level spacing of 70 and 200 K. Both approaches seem to be indistinguishable in the experimental temperature range (2–45 K).

Inelastic neutron scattering can provide a quantitative measure of the vibrational density of states of a solid as a function of energy. Although quantitative interpretation of the data from a multi-element solid is only possible with a detailed model of the lattice dynamics, qualitative information can be extracted by comparing the measured vibrational spectrum to that of a suitable reference compound. Figure 3 shows the difference in the vibrational spectrum between LaFe₄Sb₁₂ and CeFe₄Sb₁₂ obtained from neutron-scattering measurements made at the ISIS facility of the Rutherford Appleton Laboratory, UK. The scattering cross-section for Ce is much less than for La, and hence this difference spectrum tends to emphasize the features of the vibrational spectrum associated with La. The neutron data show a well defined peak at 7 meV (80 K) and a somewhat broader feature at 15 meV (175 K). These energies are in good agreement with the values obtained from the heat capacity and RUS measurements. Truly localized modes should appear as delta functions in the vibrational density of states, but hybridization with acoustic phonons will tend to broaden the peaks.

The neutron data provide strong evidence that both local modes have to be associated with the presence of the La ion, suggesting that there are two distinct eigenmodes associated with the rattling of the ion. The structure of the material makes two different La motions likely. The rattling La ion can move towards one of its nearest neighbour Sb atoms, or it can move towards a 'void'. One would expect that if the La ion moves toward a 'void', its vibrations would have a lower frequency and be more localized than if it moved towards a nearest-neighbour Sb. The neutron results are consistent with this picture. The peak at 15 meV is considerably broader than the 7-meV peak, indicating that hybridization with other modes is stronger. This implies that the 15-meV mode is less localized than the 7-meV mode. Hybridization might also be the reason why Einstein oscillators cannot perfectly describe the local modes. Higher levels of the harmonic well that are hybridized with extended phonons will lose their 'local' character, effectively reducing the harmonic oscillator to a simple TLS.

To the best of our knowledge, the only metallic compound with a well-defined local phonon mode is Al₁₀V, which has¹³ an Einstein temperature of 22 K. Compared to Al₁₀V, however, filled skutterudites have a number of advantages. Whereas Al₁₀V poses formidable materials difficulties, filled skutterudites can be made single-phase and single crystals can be grown. Moreover, by controlling the filling of the rare-earth site, the carrier concentration and thus the transport properties of the material could be modified. Further-

more, various magnetic ions could be placed on the rare-earth site, and the effects of local phonon modes on a wide variety of ground states could be studied. □

Received 2 June; accepted 6 August 1998.

1. Jeitschko, W. & Braun, D. J. LaFe₄P₁₂ with filled CoAs₃-type structure and isotopic lanthanoid-transition metal polyphosphides. *Acta Crystallogr. B* **33**, 3401–3405 (1977).
2. Braun, D. J. & Jeitschko, W. Preparation and structural investigations of antimonides with the LaFe₄P₁₂ structure. *J. Less-Common Met.* **72**, 147–156 (1988).
3. Morelli, D. M. & Meisner, G. P. Low temperature properties of the filled skutterudite CeFe₄Sb₁₂. *J. Appl. Phys.* **77**, 3777–3781 (1995).
4. Sales, B. C. *et al.* Filled skutterudite antimonides: electron crystals and phonon glasses. *Phys. Rev. B* **56**, 15081–15089 (1997).
5. Sales, B. C., Mandrus, D. & Williams, R. K. Filled skutterudite antimonides: a new class of thermoelectric materials. *Science* **272**, 1325–1328 (1996).
6. Grandjean, F. *et al.* Some physical properties of LaFe₄P₁₂ type compounds. *J. Phys. Chem. Solids* **45**, 877–886 (1984).
7. Meisner, G. P. Superconductivity in LaFe₄P₁₂. *Physica B* **108**, 763–764 (1981).
8. Meisner, G. P. *et al.* UFe₄P₁₂ and CeFe₄P₁₂: nonmetallic isotopes of superconducting LaFe₄P₁₂. *J. Appl. Phys.* **57**, 3073–3075 (1985).
9. Danebrock, M. E., Evers, C. B. H. & Jeitschko, W. Magnetic properties of alkaline earth and lanthanoid iron antimonides with the LaFe₄P₁₂ structure. *J. Phys. Chem. Solids* **57**, 381–387 (1996).
10. Feldman, J. L. & Singh, D. J. Lattice dynamics of skutterudites: first principles and model calculations for CoSb₃. *Phys. Rev. B* **53**, 6273–6282 (1996). Erratum **54**, 712 (1996).
11. Migliori, A. *et al.* Resonant ultrasound spectroscopic techniques for measurements of the elastic moduli of solids. *Physica B* **183**, 1–24 (1993).
12. Varshni, P. Temperature dependence of the elastic constants. *Phys. Rev. B* **2**, 3952–3958 (1970).
13. Caplin, D., Grüner, G. & Dunlop, J. B. Al₁₀V: An Einstein solid. *Phys. Rev. Lett.* **30**, 1138–1140 (1973).

Acknowledgements. We thank J. Feldman and R. Leisure for discussions, and A. Migliori and T. Darling for assistance with the RUS measurements. This work was supported by the Division of Materials Sciences, US Department of Energy, a Cooperative Research and Development Agreement with Marlow Industries, and the NSF.

Correspondence and requests for materials should be addressed to B.C.S. (e-mail: vb4@ornl.gov).

Synthesis of individual single-walled carbon nanotubes on patterned silicon wafers

Jing Kong^{*†}, Hyongsok T. Soh^{†‡}, Alan M. Cassell^{*}, Calvin F. Quate[‡] & Hongjie Dai^{*}

^{*} Department of Chemistry, [‡] Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA

[†] These authors contributed equally to this work

Recent progress^{1–3} in the synthesis of high-quality single-walled carbon nanotubes⁴ (SWNTs) has enabled the measurement of their physical and materials properties^{5–8}. The idea that nanotubes might be integrated with conventional microstructures to obtain new types of nanoscale devices, however, requires an ability to synthesize, isolate, manipulate and connect individual nanotubes. Here we describe a strategy for making high-quality individual SWNTs on silicon wafers patterned with micrometre-scale islands of catalytic material. We synthesize SWNTs by chemical vapour deposition of methane on the patterned substrates. Many of the synthesized nanotubes are perfect, individual SWNTs with diameters of 1–3 nm and lengths of up to tens of micrometres. The nanotubes are rooted in the islands, and are easily located, characterized and manipulated with the scanning electron microscope and atomic force microscope. Some of the SWNTs bridge two metallic islands, offering the prospect of using this approach to develop ultrafine electrical interconnects and other devices.

Our synthesis begins with the patterning of catalytic islands on silicon substrates. A schematic of the process flow is shown in Fig. 1. For regularly spaced catalytic islands on the silicon surface, the essential fabrication steps are: electron-beam lithography, deposition of Fe(NO₃)₃·9H₂O, MoO₂(acac)₂ and alumina nanoparticles in the liquid phase and lift-off (Fig. 1). The square islands are spaced at a 10-μm pitch, and the size of islands is 3 μm or 5 μm. For

chemical-vapour-deposition (CVD) growth, the patterned substrate is first heated in a tube furnace to reach 1,000 °C in an argon atmosphere. The argon flow is replaced by methane (99.99%) at a flow rate of 1,000 to 6,000 cm³ min⁻¹. The methane flow is maintained for 10 min at 1,000 °C before being replaced by argon until the furnace cools to room temperature.

We have characterized our samples with a Hitachi field emission scanning electron microscope (SEM). In Fig. 2a, b, we show two SEM micrographs recorded on the 5-μm islands. The dark, square-shaped areas are the hill-like islands. Several important features are noted in the SEM micrographs. First, we observe line-like structures coming off the islands. Some of the lines are bridging adjacent islands. Second, the lines appear dark over a brighter background of the substrate, and some of the end points of the lines can be identified. Third, the lines appear relatively straight and extend up to more than 10 μm. These SEM results indicate that carbon

nanotubes may have been synthesized from the islands. We show later that a large fraction of the nanotubes are individual SWNTs. The sole reason that we can observe individual SWNTs by SEM is because of the electrical connections existing between nanotubes and islands which render them as dark lines in the SEM images.

Further sample characterization was carried out using an atomic force microscope (AFM). An AFM image recorded on 5-μm islands,

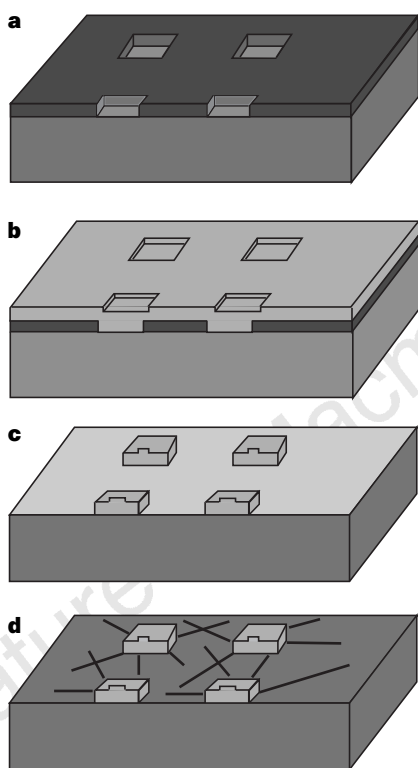


Figure 1 Schematic of process flow. **a–c**, Process flow for fabricating catalytic islands on silicon substrates used in CVD nanotube growth (**d**). **a**, Electron beam lithography was used to fabricate square holes in a polymethylmethacrylate (PMMA) film on silicon. We used p-type Si(100) substrate (resistivity 10 Ω cm) without removing the native oxide. The 0.25-μm-thick PMMA film was obtained by spin coating (at 4,000 r.p.m.) the silicon substrate with a 5% PMMA (relative molecular mass, 495K) chlorobenzene solution, followed by baking at 170 °C for 2 h. Electron-beam patterning was done using a Hitachi HL700F machine operated at 30 kV; exposure dosage was 500 μC cm⁻². The exposed PMMA was removed with a methyl isobutyl ketone (MIBK) and isopropyl alcohol (3:1) solution. **b**, 0.05 mmol Fe(NO₃)₃·9H₂O, 0.015 mmol of MoO₃ (acac)₃ and 15 mg alumina nanoparticles (Degussa, aluminum oxide C; average particle size, 14 nm) was added to 15 ml methanol. The mixture was stirred for 24 h and sonicated for 1 h to afford a 0.05 mmol/15 mg/15 ml Fe(NO₃)₃·9H₂O/alumina/methanol suspension. A drop of the suspension was deposited onto the patterned PMMA substrate. After solvent vaporization at room temperature, the substrate was heated at 170 °C for 5 min. **c**, Lift-off of PMMA in 1,2-dichloroethane leads to the final substrate containing catalyst islands. **d**, CVD of methane at 1,000 °C produces SWNTs off the islands (see other figures and text).

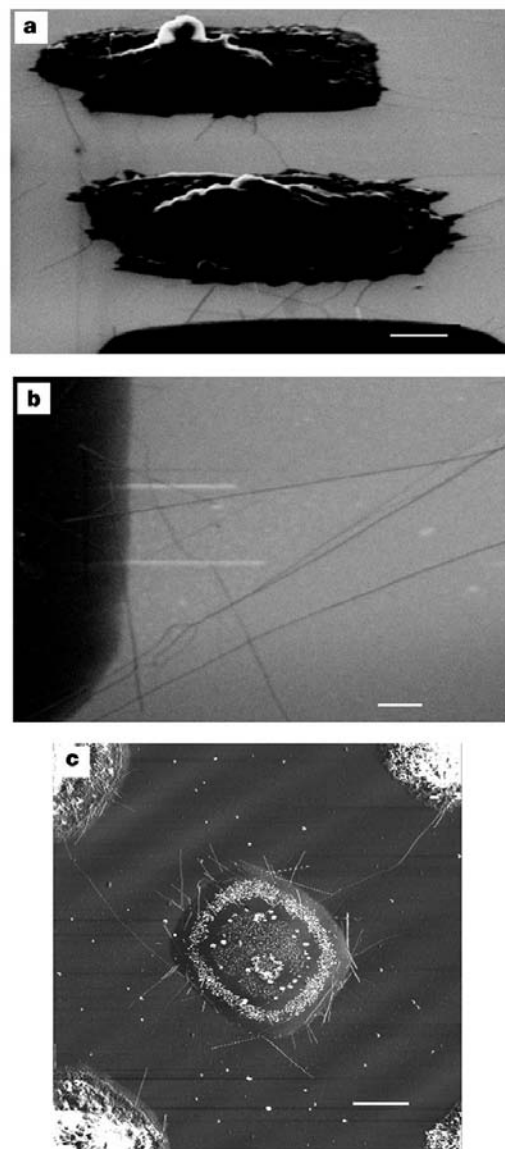


Figure 2 Nanotubes grown on a catalyst-patterned substrate. **a**, Scanning electron microscope (SEM) image that shows line-like structures (nanotubes; see text) extending off the patterned 5-μm islands on silicon after methane CVD. The sample was tilted at a 65° angle relative to the incident electron beam (25 kV). The dark protrusions on the surface are the square islands. Nanotubes appear as dark lines because they have electrical connections with the metallic islands. The brighter background is due to the electron charging effect of the silicon oxide substrate. Note that some of the nanotubes are bridging islands. Scale bar: 1 μm. **b**, A high-magnification (scale bar: 300 nm) SEM image recorded with the same sample as in **a**. Clearly, nanotubes are connected to the island (dark region) and most of the nanotubes appear extremely straight at the micrometre scale. **c**, Typical large-scale (scale bar: 2 μm) phase image recorded by tapping mode AFM, showing carbon nanotubes grown from the patterned islands and bridging between islands.

shown in Fig. 2c, reveals similar line-like structures. Extensive topographic imaging by tapping mode AFM found that the line-like structures are indeed synthesized nanotubes with higher topography than the Si substrate. The topography image in Fig. 3a shows that all nanotubes are connected with the islands. Several nanotubes have one end resting in between islands, whereas nanotubes that bridge islands have both ends buried in the two opposing islands. From height measurements, we determined that the diameters of the nanotubes in Fig. 3a range from 0.8 nm to 3.0 nm, which strongly suggests that individual SWNTs were synthesized off the islands. In Fig. 3b, we show an SWNT (2.2 nm in diameter) bridging two diagonal islands. We found that lowering the concentrations of $\text{Fe}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$ and alumina led to the formation of islands shaped like doughnuts. Figure 3c, d shows AFM images recorded on 3- μm doughnut-shaped islands. The

average diameter of nanotubes in Fig. 3c is 1.6 nm, and the length of nanotubes extends up to more than 20 μm . To illustrate further the high quality of the synthesized individual SWNTs, we show in Fig. 3d a higher-magnification AFM topography image. Note that nanotubes on the line-cut (Fig. 3d) have diameters that range from 1.2 to 2.7 nm (Fig. 3e).

We emphasize several important features of the nanotube materials synthesized by this approach. First, about 90% of the nanotubes, having diameters in the range of 1–3 nm, are individual SWNTs. SWNTs with diameters ranging from ~ 1 up to 5 nm have been observed previously in materials synthesized by arc discharge^{9,10}. Second, we found that the SWNTs are remarkably straight, considering their extreme aspect ratios of up to 10^4 . So far we have not observed nanotubes with frequent kinks along their length: only smooth curving exists. These SWNTs must be nearly free of topological defects arising from pentagons and heptagons along the side walls. Third, our synthetic method is very simple and can be scaled up on full-size silicon wafers. Finally, our synthesis directly fabricates high-quality individual SWNT devices needed for fundamental and practical purposes¹¹. Essentially, the frequently observed nanotube bridges crosslink the micrometre-scale islands

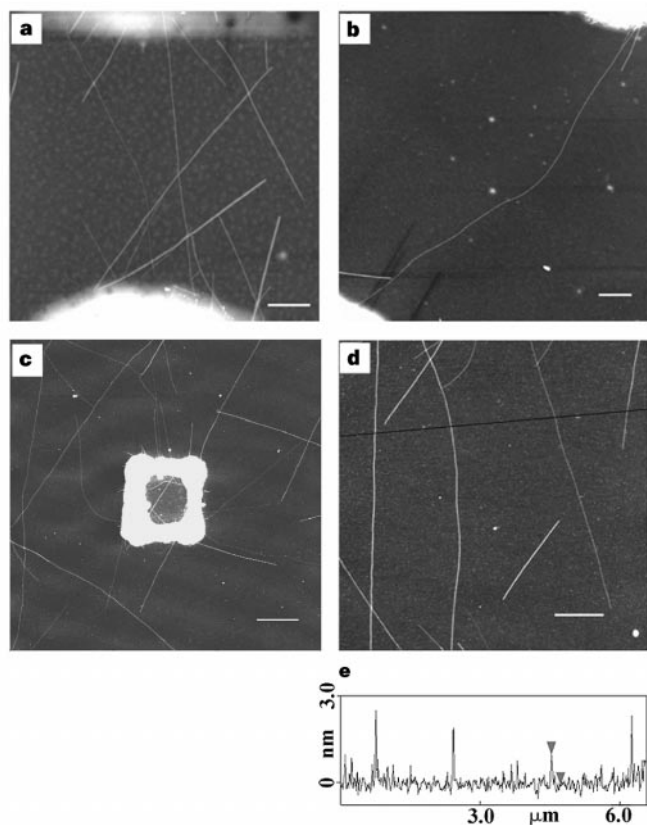


Figure 3 Topography images of individual SWNTs recorded by tapping mode AFM. Commercial silicon cantilevers with $\sim 100 \text{ N m}^{-1}$ stiffness and $\sim 250 \text{ kHz}$ resonance frequency were used. **a**, Image recorded on the 5- μm islands showing individual SWNTs synthesized off the islands. The bright regions at the top and bottom of the image are the edges of two adjacent islands. Nanotubes in this image have diameters between 0.9 and 3 nm (scale bar: 500 nm). **b**, Image recorded with the 5- μm islands showing a 2-nm diameter individual SWNT bridging two islands. The bright regions at the upper right and lower-left corners of the image are the edges of two diagonal islands (scale bar: 500 nm). **c**, Data recorded with a 3- μm doughnut-shaped island showing long individual SWNTs growing off doughnut-shaped islands and making a network. In the image, several nanotubes are from adjacent islands, the average nanotube diameter is 1.6 nm, and the nanotubes are up to 20 μm long (scale bar: 2 μm). **d**, Image of individual SWNTs showing that each nanotube appears uniform in both width and height (scale bar: 1 μm). We note a mysterious nanotube in this image that has both ends unconnected with any island, a phenomenon that occurs rarely. **e**, Topographic height measurements for nanotubes on the line-cut (black line) in **d**. The measured diameters of the four individual SWNTs are 2.7, 1.8, 1.2 and 1.8 nm, respectively.

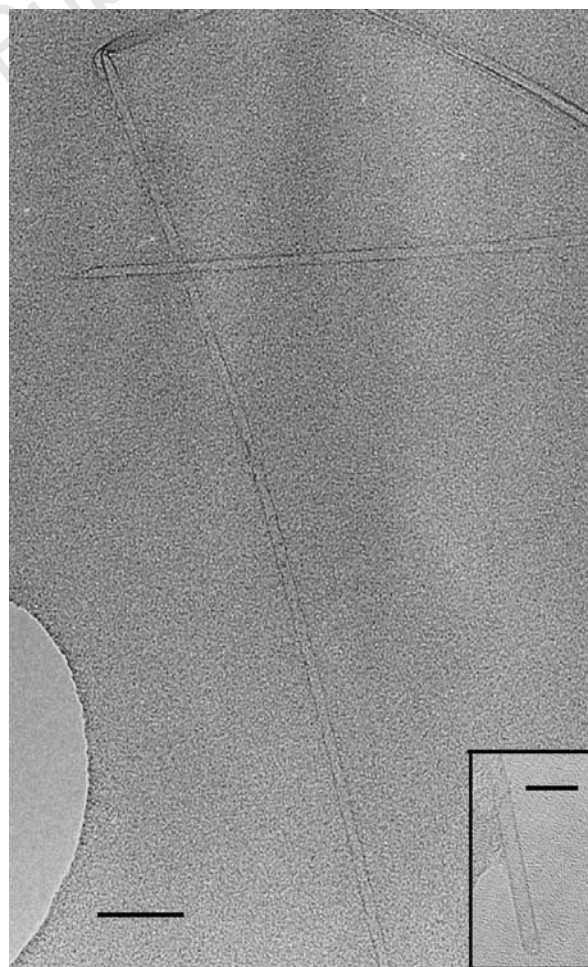


Figure 4 A transmission electron microscope (TEM) image (scale bar: 50 nm) of an individual SWNT synthesized on supported catalysts by the methane CVD approach. The image background is due to the holey carbon film on the TEM grid. The inset (scale bar: 10 nm) shows a typical closed particle-free end of an individual nanotube. We note that a small fraction of the SWNTs are in bundled form (for detailed results, see ref. 17). For TEM sample preparation, we suspended 10 mg of the synthesized material in methanol by sonication for 15 min, and then put a few drops of the suspension onto the TEM grid.

into a macroscopic electrical network. Thus, electrical connections from the macroscopic scale to individual SWNTs can now be easily made via the micrometre-scale islands. Using this approach, we have designed and fabricated individual SWNT electrical circuits by direct CVD syntheses on insulating SiO₂ substrates with the patterned catalyst islands connected to macroscopic metal electrodes (see Supplementary Information). By mechanical manipulation of nanotube bridges using contact mode AFM¹², we have obtained evidence of strong mechanical fixation of nanotubes to the islands. Further, we have measured the lateral forces needed to break individual SWNTs and estimated the tensile strength of SWNTs (of the order of 200 GPa; see Supplementary Information).

We have chosen methane for carbon feedstock in our CVD synthesis because methane is the most kinetically stable hydrocarbon at elevated temperatures. Previously, extensive work has been done using methane in catalysed CVD synthesis of carbon fibres at temperatures up to 1,100 °C (refs 13–16). At 1,000 °C, the catalytic decomposition of methane by our catalyst dominates over self-pyrolysis, at least within a timescale of 10 min. This is evident from our observation that no obvious amorphous carbon coating exists on the side walls of the nanotubes. However, an amorphous carbon coating can build up over time. Previous work on micrometre-scale carbon-fibre synthesis showed that as the reaction is allowed to take place for hours, methane pyrolysis at ~1,000 °C leads to an appreciable thickness of carbon coating over a nanotube core^{13–16}. By limiting our CVD synthesis to 10 min and using a relatively high flow rate, we are able to avoid the amorphous carbon coating.

To confirm that the methane CVD approach produces individual SWNTs, we prepared bulk catalyst supported on alumina nanoparticles. The catalyst preparation involves impregnation of 4.0 g alumina nanoparticles with 1.0 g Fe(NO₃)₃·9H₂O in 30 ml methanol for 24 h, after which the methanol is removed by evaporation at 80 °C. Using this bulk catalyst, we carried out methane CVD under exactly the same conditions as with the patterned Si substrates. Using a transmission electron microscope (TEM), we found abundant SWNTs (see Fig. 4 and ref. 17). The diameter range of the individual SWNTs was 1–5 nm, with a mean of 1–2 nm. Ends of SWNTs were frequently observed under TEM. All of the ends imaged so far appear to be closed and free of metal particles (Fig. 4, inset). From these observations, we infer that the SWNTs grow in methane CVD by a 'base-growth' mechanism. Base growth and tip growth are the two mechanisms that have been proposed for CVD growth of various forms of carbon tubes^{18–21}. The first step of such a CVD process involves the absorption and decomposition of hydrocarbon molecules on the surface of a transition metal (Fe, Ni, Co, for example) catalyst, followed by diffusion of carbon atoms into the catalyst bulk from the supersaturated surface^{19–21}. The tip-growth model involves a metal catalyst particle at a nanotube end being carried away as the nanotube lengthens: thus, the carried-along particle is responsible for supplying carbon feedstock needed for the nanotube growth. In contrast, base growth involves a nanotube lengthening with a particle-free closed end. Carbon feedstock is supplied from the base where the nanotube interfaces with the catalyst material. In our methane CVD system, because there are only closed particle-free nanotube ends we conclude that the base-growth mechanism operates under our conditions. For SWNT syntheses on patterned substrates, we are effectively making catalyst *in situ* in PMMA-defined Petri dishes (Fig. 1). Our synthetic strategy should allow rational design of various single-walled nanotube devices by direct chemical synthesis. □

Received 6 April; accepted 7 July 1998.

1. Thess, A. *et al.* Crystalline ropes of metallic carbon nanotubes. *Science* **273**, 483–487 (1996).
2. Journet, C. *et al.* Large-scale production of single-walled carbon nanotubes by the electric-arc technique. *Nature* **388**, 756–758 (1997).
3. Bethune, D. S. *et al.* Cobalt-catalysed growth of carbon nanotubes with single-atomic-layer walls. *Nature* **363**, 605–607 (1993).
4. Iijima, S. & Ichihashi, T. Single-shell carbon nanotubes of 1-nm diameter. *Nature* **363**, 603–605 (1993).

5. Tans, S. J. *et al.* Individual single-wall carbon nanotubes as quantum wires. *Nature* **386**, 474–477 (1997).
6. Bockrath, M. *et al.* Single-electron transport in ropes of carbon nanotubes. *Science* **275**, 1922–1925 (1997).
7. Odom, T., Huang, J., Kim, P. & Lieber, C. M. Atomic structure and electronic properties of single-walled carbon nanotubes. *Nature* **391**, 62–64 (1998).
8. Wildoer, J. W. G., Venema, L. C., Rinzler, A. G., Smalley, R. E. & Dekker, C. Electronic structure of atomically resolved carbon nanotubes. *Nature* **391**, 59–62 (1997).
9. Kiang, C.-H., Goddard, W. A., Beyers, R., Salem, J. R. & Bethune, D. S. Catalytic synthesis of single-layer carbon nanotubes with a wide range of diameters. *J. Phys. Chem.* **98**, 6612–6618 (1994).
10. Lin, X., Wang, X. K., Dravid, V. P., Chang, R. P. H. & Ketterson, J. B. Large scale synthesis of single-shell carbon nanotubes. *Appl. Phys. Lett.* **64**, 181–183 (1994).
11. Iijima, S. Carbon nanotubes. *MRS Bull.* **19**, 43–49 (1994).
12. Wong, E. W., Sheehan, P. E. & Lieber, C. M. Nanobeam mechanics—elasticity, strength, and toughness of nanorods and nanotubes. *Science* **277**, 1971–1975 (1997).
13. Tibbetts, G. G. in *Carbon Fibers, Filaments and Composites* 73–94 (Kluwer Academic, Amsterdam, 1990).
14. Tibbetts, G. G. Carbon fibers produced by pyrolysis of natural gas in stainless steel tubes. *Appl. Phys. Lett.* **42**, 666–668 (1983).
15. Jaeger, H. & Behrsing, T. The dual nature of vapour-grown carbon fibres. *Composites Sci. Technol.* **51**, 231–242 (1994).
16. Qin, L. C. & Iijima, S. Fibrilliform growth of carbon nanotubes. *Mater. Lett.* **30**, 311–314 (1997).
17. Kong, J., Cassell, A. & Dai, H. Chemical vapor deposition of methane for single-walled carbon nanotubes. *Chem. Phys. Lett.* **292**, 567–574 (1998).
18. Amelinckx, S. *et al.* A formation mechanism for catalytically grown helix-shaped graphite nanotubes. *Science* **265**, 635–639 (1994).
19. Baker, R. T. K. Catalytic growth of carbon filaments. *Carbon* **27**, 315–323 (1989).
20. Tibbetts, G. G., Devour, M. G. & Rodda, E. J. An adsorption-diffusion isotherm and its application to the growth of carbon filaments on iron catalyst particles. *Carbon* **25**, 367–375 (1987).
21. Tibbetts, G. G. Why are carbon filaments tubular? *J. Cryst. Growth* **66**, 632–638 (1984).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank J. Brauman, J. Han, C. Marcus, T. Kenny, A. Kapitulnik and A. Morpurgo for helpful discussions, and J. Kim for his assistance with SEM. This work is partly supported by a Camille and Henry Dreyfus New Faculty Award (H.D.); the NSF and the Office of Naval Research (JSEP) (C.F.Q.); and NASA Ames Research Center (A.M.C.).

Correspondence and requests for materials should be addressed to H.D. (e-mail: hdoi@chem.stanford.edu).

Energy implications of future stabilization of atmospheric CO₂ content

Martin I. Hoffert*, Ken Caldeira†, Atul K. Jain‡, Erik F. Haites§, L. D. Danny Harvey||, Seth D. Potter*¶, Michael E. Schlesinger‡, Stephen H. Schneider#, Robert G. Watts*, Tom M. L. Wigley** & Donald J. Wuebbles‡

* Department of Physics, New York University, 4 Washington Place, New York, New York 10003-6621, USA

† Lawrence Livermore National Laboratory, Livermore, California 94550, USA

‡ Department of Atmospheric Sciences, University of Illinois, Urbana, Illinois 61801, USA

§ Margaree Consultants, Toronto, M5H 2X6, Canada

|| Department of Geography, University of Toronto, Toronto, M5S 3G3, Canada

Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

¶ Department of Mechanical Engineering, Tulane University, New Orleans, Louisiana 70118, USA

** National Center for Atmospheric Research, Boulder, Colorado 80307, USA

The United Nations Framework Convention on Climate Change¹ calls for "stabilization of greenhouse-gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system...". A standard baseline scenario^{2,3} that assumes no policy intervention to limit greenhouse-gas emissions has 10 TW (10 × 10¹² watts) of carbon-emission-free power being produced by the year 2050, equivalent to the power provided by all today's energy sources combined. Here we employ a carbon-cycle/energy model to

¶ Present address: Boeing, Seal Beach, California 90740-7644, USA.

estimate the carbon-emission-free power needed for various atmospheric CO₂ stabilization scenarios. We find that CO₂ stabilization with continued economic growth will require innovative, cost-effective and carbon-emission-free technologies that can provide additional tens of terawatts of primary power in the coming decades, and certainly by the middle of the twenty-first century, even with sustained improvement in the economic productivity of primary energy. At progressively lower atmospheric CO₂-stabilization targets in the 750–350 p.p.m.v. range, implementing stabilization will become even more challenging because of the increasing demand for carbon-emission-free power. The magnitude of the implied infrastructure transition suggests the need for massive investments in innovative energy research.

Intergovernmental Panel on Climate Change (IPCC) working groups developed six scenarios for greenhouse-gas emissions based on socioeconomic projections^{2,3}. Scenario IS92a incorporated widely accepted projections and the then-current consensus on population, economic development and energy technology to generate projections of greenhouse-gas emissions for the twenty-first century assuming “no new climate change policies”. This baseline IS92a scenario has been dubbed ‘business as usual’.

In general, the rate at which carbon is emitted (as CO₂) by energy production is given by the Kaya identity^{4–6},

$$\dot{M}_c = N(\text{GDP}/N)(\dot{E}/\text{GDP})(C/E) \quad (1)$$

expressing emissions as the product of population (N), per capita gross domestic product (GDP/N), primary energy intensity ($\dot{E}/\text{GDP}$) and carbon intensity (C/E). Here we express the rate of primary energy consumption from all fuel sources (the “burn rate”) in watts (W) and the gross domestic product in (1990 US) \$ yr⁻¹ so their ratio, the energy intensity, has units of Wyr \$⁻¹. Carbon intensity, the weighted average of the carbon-to-energy emission factors of all energy sources, has the units kg C W⁻¹ yr⁻¹. For example, from equation (1) fossil-fuel CO₂ emissions in 1990 were $\dot{M}_c = 5.3 \times 10^9$ persons $\times$ 4,100 \$ per person yr⁻¹ $\times$ 0.49 Wyr \$⁻¹ $\times$ 0.56 kg C W⁻¹ yr⁻¹ $\approx$ 6.0 Gt C yr⁻¹ (1 Gt C = 10¹² kg C).

To illustrate the relative contributions of the factors in equation (1) historically and under IS92a, we evaluated each of them globally over the 210-year period from 1890 to 2100 (Fig. 1) from historical data^{7–9} before 1990, and from documents defining IPCC scenarios^{2,10–12} after 1990. Although some of this information is implicit in IPCC documents, it has not been previously presented in this way.

Although illustrating the ‘big picture’, aggregating Kaya decomposition terms globally can mask regional developmental differences¹³. Data for individual nations indicate that $\dot{E}/\text{GDP}$ typically increases during economic development, reaching a maximum as infrastructure investments peak, and declining only after some lag as economic productivity rises and the economy shifts structurally to less-energy-intensive activities (for example,

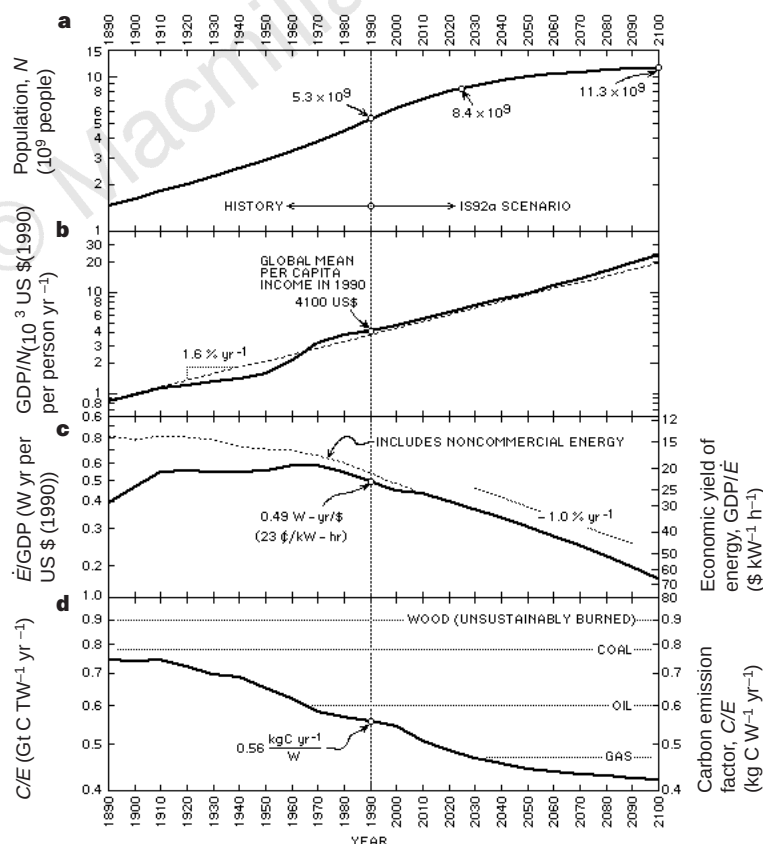


Figure 1 Evolution of factors governing the rate of global fossil-fuel carbon emissions in the Kaya identity: $\dot{M}_c = N(\text{GDP}/N)(\dot{E}/\text{GDP})(C/E)$. Historical curves (1890–1990) are from archival data^{7–9}; future projections (1990–2100) are computed for the IPCC IS92a scenario^{2,10–12}. GDP is inflation-corrected to 1990 US dollars. **a**, Global population; **b**, Per capita GDP; **c**, primary energy intensity ($\dot{E}/\text{GDP}$: left hand scale) and economic productivity of energy ($\text{GDP}/\dot{E}$: right hand scale); **d**, carbon intensity of the energy mix; the horizontal lines are emission factors of

individual carbonaceous fuels. Global population, N , shown after 1990 is the UN mid-range projection made at the time the IS92 scenarios were developed. It reaches 11.3 billion by 2100. Per capita global mean GDP continues its monotonic rise at $\sim 1.6\%$ yr⁻¹ over the entire twenty-first century. The primary energy intensity incorporates both engineering energy efficiency and structural changes in the economy governing the material content of economic growth.

services)¹⁴. Apparently declining, energy intensities of India and China today are still 2 to 5 times the global mean⁸. To focus on energy supply issues we provisionally accept the IS92a projections of 1% yr⁻¹ improvement in global $\dot{E}/\text{GDP}$, recognizing that achieving this will depend crucially on the technology and structural changes adopted by developing nations¹⁵.

Another opportunity for emission reductions is continuation of the “decarbonization” of the past 100 years reflected in decreasing carbon intensity of the global energy mix⁸. Under IS92a, the global mean C/E continues to decrease monotonically over the next century (Fig. 1d). Indeed, the evolving global energy mix based on assumed declining costs of nuclear and carbon-free energy relative to fossil fuels has global C/E dropping to that of natural gas by 2030; and it declines even more thereafter. Such rapid decarbonization is possible only by the massive introduction of carbon-free power, ~10 TW by 2050. Even with this much carbon-free power and sustained 1% yr⁻¹ improvements in energy intensity, the net effect of the factors in equation (1) more than doubles 1990 CO₂ emissions by 2050.

Figure 2 shows carbon emissions (Fig. 2a), primary power levels (Fig. 2b) and carbon-free primary power (Fig. 2c) required over the twenty-first century for IS92a and CO₂ stabilization scenarios. Even

the optimistic decline of the last two factors in the Kaya identity cannot prevent emissions from increasing from 6 Gt C yr⁻¹ in 1990 to ~20 Gt C yr⁻¹ by 2100 under ‘business as usual’ (Fig. 2a). Also shown as differently shaded zones are the relative contributions of natural gas, oil and coal to emissions. A feature of IS92a worth noting is that the share of carbon-intensive coal, relative to less-carbon-intensive natural gas and oil, rises after 2025, but the carbon intensity (C/E) of the fuel mix declines overall, a feature possible only by the massive introduction of carbon-free energy sources.

The curves in Fig. 2a are allowable emission levels over time which ultimately stabilize atmospheric CO₂ at 750, 650, 550, 450 and 350 p.p.m.v. computed for the Wigley, Richels and Edmonds stabilization paths^{16,17}. These delayed stabilization paths, which follow IS92a emissions early on, with large reductions later (hereafter, WRE 350, 450, ..., 750 scenarios), could buy time to attain CO₂ stabilization goals. This is possible because a given atmospheric CO₂ concentration goal depends roughly on cumulative carbon emissions and can be approached by an infinite number of paths, some of which constrain emissions early, some later. However, Wigley, Richels and Edmonds did not consider whether their paths were a realistic transition from the present fossil-fuel system. They emphasized that their “results should not be interpreted as a ‘do nothing’ or ‘wait and see’ policy”, calling for prompt and sustained commitment to research, development and demonstration to ensure that low-carbon and carbon-free energy alternatives are available when needed¹⁶.

The black topmost curve in Fig. 2b shows the evolution of total primary power required to meet the economic goals of IS92a, with gas, oil, coal, nuclear and renewable components shown as coloured areas. Also shown are allowable primary power levels from fossil-fuel sources computed for WRE stabilization scenarios. The difference between the IS92a total primary power, $\dot{E}$, and fossil-fuel power allowable for CO₂ stabilization, $\dot{E}_f$, must be provided by carbon-free sources if the economic and “efficiency” assumptions of IPCC ‘business as usual’ are maintained; an increasingly challenging goal as the CO₂ concentration target is lowered.

Figure 2c shows carbon-free power, $\dot{E}_{cf} = \dot{E} - \dot{E}_f$, required for IS92a and for CO₂ stabilization via WRE 350–750 paths in a world

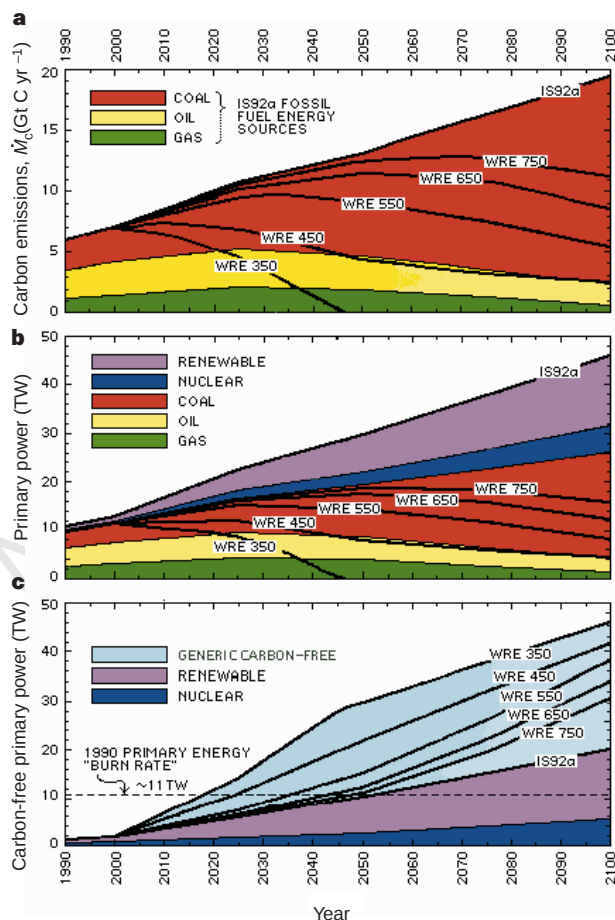


Figure 2 Fossil-fuel carbon emissions and primary power in the twenty-first century for IPCC IS92a and WRE stabilization scenarios. **a**, Carbon emissions; **b**, primary power and **c**, carbon-free primary power. Coloured areas are gas, oil, coal, nuclear and renewable components of IS92a from the energy economics model of Pepper *et al.*¹². Carbon emissions for WRE scenarios¹⁶ are outputs of our inverse carbon-cycle model¹⁷. Fossil-fuel power for WRE scenarios is based on IS92a burn rates of gas, oil and coal employed sequentially in descending priority, and limited by total carbon emission caps. Carbon-free primary power is total primary power less fossil-fuel carbon power.

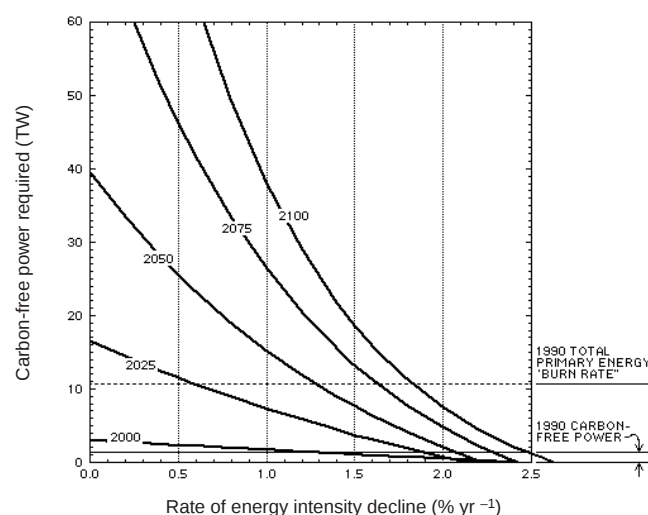


Figure 3 Twenty-first century trade-offs, between carbon-free power required and “energy efficiency”, to stabilize at twice the pre-industrial CO₂ concentration. Carbon-free primary power is plotted versus the annual rate of decline in energy intensity. The latter is defined as $-\left[\frac{d(\dot{E}/\text{GDP})}{dt}\right](\dot{E}/\text{GDP})^{-1}100\%$. All cases assume WRE 550 CO₂ stabilization paths and GDP growth of 2.9% yr⁻¹ to 2025, 2.3% yr⁻¹ thereafter. The rate of energy intensity decline in IS92a and in our CO₂ stabilization base cases is ~1.0% yr⁻¹.

economy growing as IS92a and with the same improvement rate in $\dot{E}/\text{GDP}$. In that case stabilizing CO_2 concentrations at 1990 levels according to WRE 350 will require ~ 10 TW of carbon-free primary power by 2018, equal to the total 1990 primary power of the world economy. WRE 550, which leads to an approximate doubling of pre-industrial atmospheric CO_2 , requires this much carbon-free power by 2035. These results imply a massive transition from the present fossil-fuel-dominated energy infrastructure to carbon-free sources in the coming decades to stabilize CO_2 at reasonable target values. Even IS92a requires ~ 10 TW carbon-free power by 2050.

Studies by US DOE laboratories in support of the Kyoto negotiations to reduce carbon emissions in the 1998–2010 time frame emphasize demand side reductions from improved energy efficiency in motor vehicles, buildings and electrical appliances¹⁸. Our analysis indicates that beyond 2010 new carbon-free primary power technologies will increasingly be needed even with significant improvements in the ability to convert primary power into GDP.

Figure 3 shows the trade-off between increases in carbon-free power and energy efficiency improvements, expressed as the annual percentage improvement in $\dot{E}/\text{GDP}$ required to achieve a $2 \times$ pre-industrial CO_2 goal by the WRE 550 path. Sustained improvement of energy intensity of order $\pm 1.0\% \text{ yr}^{-1}$ relative to the base case ($1\% \text{ yr}^{-1}$) creates increasingly large differences in carbon-free power required as the twenty-first century progresses (Fig. 3). For a $2\% \text{ yr}^{-1}$ compounded growth, the carbon-free power required remains modest even by the year 2100. But $2\% \text{ yr}^{-1}$ may be almost impossible to sustain over the next century, as that would imply growing the world population and economy significantly at nearly constant primary power. If there is no improvement in energy intensity, some 40 carbon-free terawatts could be needed by 2050.

The authors of the IPCC central 'business as usual' scenario (IS92a) believed that an exponential decline in $\dot{E}/\text{GDP}$ of $1\% \text{ yr}^{-1}$ would be sustainable over the next century employing only those emission-control policies internationally agreed to at the 1992 Rio climate treaty negotiations¹². But either deploying 10–30 TW carbon-free power or improving $\dot{E}/\text{GDP}$ by $2\% \text{ yr}^{-1}$ will be very difficult. The need to push hard on both improving energy supply and reducing demand is demonstrated by our analysis, as well as the fact that there are real trade-offs—more of one can significantly diminish the need for the other.

Stabilizing atmospheric CO_2 at twice pre-industrial levels while meeting the economic assumptions of 'business as usual' implies a massive transition to carbon-free power, particular in developing nations. There are no energy systems technologically ready at present to produce the required amounts of carbon-free power⁶. Some suggest the answer may be integrated energy systems based on fossil fuels¹⁹ in which carbon dioxide²⁰ or solid carbon²¹ is sequestered in reservoirs isolated from the atmosphere, or "geoengineering" compensatory climate changes^{22–24}. Despite potentially serious environmental and cost problems, these approaches would allow fossil fuel, increasingly coal, to continue its historic rise as the primary energy source of the next century.

It is time now to look hard at the engineering feasibility of transformative technologies that can change the way primary energy itself is produced. It is within the range of climate change and impact projections that stabilization of atmospheric CO_2 at some level below the IS92a baseline is necessary to mitigate large adverse effects on global economies and ecosystems³. In that case, a massive infusion of new energy-producing technologies at the required scale could be needed to prevent "dangerous anthropogenic interference with the climate system" (ref. 1).

Analyses of carbon emissions targets thus far have quite reasonably emphasized market economics theory. Some suggest that market forces alone, perhaps supplemented by carbon taxes, are sufficient to stimulate adequate levels of innovation in emission-reducing technologies. However, market inefficiencies may preclude timely development of such technologies at the required scale²⁵.

There are renewable^{26,27}, fission²⁸ and fusion²⁹ concepts incorporating innovative technological ideas at early research and development stages that could, in principle, provide the needed carbon-free power. But without policy incentives to overcome socioeconomic inertia these could take more than 50 years to penetrate to their market potential³⁰.

These results underscore the pitfalls of 'wait-and-see'. This past century, accelerated technology development from wartime and postwar research produced commercial aviation, radar, computer chips, lasers and the Internet, among other things. Researching, developing and commercializing carbon-free primary power technologies capable of 10–30 TW by the mid-twenty-first century could require efforts, perhaps international, pursued with the urgency of the Manhattan Project or the Apollo space programme. The roles of governments and market entrepreneurs in the eventual deployment of such technologies need to be considered more comprehensively than we can do here. But the potentially adverse effect of humanity on the Earth's climate could well stimulate new industries in the twenty-first century, as did the Second World War and the 'cold war' in this century. □

Received 10 August; accepted 8 October 1998.

1. United Nations Framework Convention on Climate Change (Text) (UNEP/WMO, Climate Change Secretariat, Geneva, 1992).
2. Leggett, J. et al. in *Climate Change 1992: the Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T. et al.) 69–95 (Cambridge Univ. Press, 1992).
3. Houghton, J. T. et al. (eds) *Climate Change 1995: the Science of Climate Change* (Cambridge Univ. Press, 1996).
4. Kaya, Y. Impact of carbon dioxide emission control on GNP growth: Interpretation of proposed scenarios (IPCC Response Strategies Working Group Memorandum, 1989).
5. Nakićenović, N., Victor, D., Grübler, A. & Schrattenholzer, L. Long term strategies for mitigating global warming: Introduction. *Energy* **18**, 403–409 (1993).
6. Hoffert, M. I. & Potter, S. D. in *Engineering Response to Global Climate Change* (ed. Watts, R. G.) 205–260 (Lewis, Boca Raton, FL, 1997).
7. de Vries, B. & van den Wijngaart, R. *The Targets/Image Energy (TIME) 1.0 Model* (GLOBO Rep. Ser. No. 16, National Institute of Public Health and the Environment (RIVM), Bilthoven, Netherlands, 1995).
8. Nakićenović, N. Freeing energy from carbon. *Daedalus* **125**(3), 95–112 (1996).
9. Boden, T. A. et al. (eds) *Trends 93: A Compendium of Data on Global Change* 501–584 (ORNL/CDIA-65, Carbon Dioxide Information Analysis Center, Oak Ridge National Lab., Oak Ridge, TN, 1994).
10. Alcamo, J. et al. in *Climate Change 1994* (eds Houghton, J. T. et al.) 247–304 (Cambridge Univ. Press, Cambridge, UK, 1995).
11. Alcamo, J. (ed.) *Image 2.0: Integrated Modeling of Global Climate Change* (Kluwer, Netherlands, 1994).
12. Pepper, W. J. et al. *Emission Scenarios for the IPCC—An Update* (US Environmental Protection Agency, Washington DC, 1992).
13. Yang, C. & Schneider, S. H. Global carbon dioxide scenarios: sensitivity to social and technological factors in three regions. *Mitigat. Adapt. Strat. for Global Change* **2**, 373–404 (1998).
14. Reddy, A. K. N. & Goldemberg, J. in *Energy for Planet Earth* (ed. Piel, J.) 59–71 (Freeman, New York, 1991).
15. Goldemberg, J. Energy needs in developing countries and sustainability. *Science* **269**, 1058–1059 (1995).
16. Wigley, T. M. L., Richels, R. & Edmonds, J. A. Economic and environmental choices in the stabilization of atmospheric CO_2 concentration. *Nature* **379**, 240–243 (1996).
17. Jain, A. K., Khesghi, H. S., Hoffert, M. I. & Wuebbles, D. J. Distribution of radiocarbon as a test of global carbon cycle models. *Glob. Biogeochem. Cycles* **9**, 153–166 (1995).
18. Romm, J., Levine, M., Brown, M. & Peterson, E. A road map for U.S. carbon reductions. *Science* **279**, 669–670 (1998).
19. National Academy of Sciences *Policy Implications of Greenhouse Warming* 340–375 (National Academy Press, Washington DC, 1992).
20. Parson, E. A. & Keith, D. W. Fossil fuels without CO_2 emissions: Progress, prospects and policy implications. *Science* (in the press).
21. Steinberg, M. & Dong, Y. in *Global Climate Change: Science, Policy and Mitigation Strategies* (eds Mathai, C. V. & Stensland, C.) 858–873 (Air & Waste Management Assoc., Pittsburgh, PA, 1994).
22. Marland, G. (ed.) Special section on geoengineering. *Clim. Change* **33**(3), 275–336 (1996).
23. Flannery, B. P. et al. in *Engineering Response to Global Climate Change* (ed. Watts, R. G.) 379–427 (Lewis, Boca Raton, FL, 1997).
24. Fogg, M. J. *Terraforming: Engineering Planetary Environments* (SAE International, Warrendale, PA, 1995).
25. Schneider, S. H. & Goulder, L. H. Achieving low-cost emission targets. *Nature* **389**, 13–14 (1997).
26. Johansson, T. M. et al. (eds) *Renewable Energy* (Island, Washington DC, 1993).
27. Glaser, P. E. et al. (eds) *Solar Power Satellites: A Space Energy System for Earth* (Wiley-Praxis, Chichester, 1998).
28. Weinberg, A. M. in *Technologies for a Greenhouse-Constrained Society* (eds Kuliasha, M. A. et al.) 227–237 (Lewis, Boca Raton, FL, 1992).
29. Fowler, T. K. *The Fusion Quest* (Johns Hopkins, Baltimore, 1997).
30. Grübler, A. in *Technological Trajectories and the Human Environment* (eds Ausubel, J. H. & Langford, H. D.) 14–32 (National Academy Press, Washington DC, 1997).

Acknowledgements. We thank DOE, NASA and NSF for partial support of this work. We also thank the Aspen Global Change Institute for discussions during the 1998 summer workshop 'Innovative Energy Systems and CO_2 Stabilization'.

Correspondence and requests for materials should be addressed to M.I.H. (e-mail: marty.haffert@nyu.edu).

Oil preserved in fluid inclusions in Archaean sandstones

Adriana Dutkiewicz*, Birger Rasmussen† & Roger Buick‡

* CSIRO Division of Petroleum Resources, PO Box 136, North Ryde New South Wales 1670, Australia

† Centre for Strategic Mineral Deposits, Department of Geology & Geophysics, University of Western Australia, Nedlands Western Australia 6907, Australia

‡ School of Geosciences, University of Sydney, Sydney New South Wales 2006, Australia

Oil is generally thought to be geologically young, as it is thermodynamically unstable when subjected to elevated temperatures over long periods in open systems^{1,2}. Indeed, almost all petroleum production comes from rocks younger than 400 million years (ref. 3). Although the oldest known oil occurs in rocks 1,650 Myr old⁴, suitable source rocks were abundant in older geological successions⁵ and circumstantial evidence suggests that some of these generated hydrocarbons early in their history⁶. Here, we report the discovery of oil preserved in fluid inclusions in sandstones dating back ~3,000 Myr. Most inclusions lie within healed microfractures confined to individual detrital quartz grains, indicating that their oil was emplaced before Archaean or Palaeoproterozoic metamorphism sealed all voids and thus came from older (in some cases Archaean) sources. The fluid inclusions apparently acted as inert pressure vessels that protected the oil from subsequent degradation by circulating fluids and mineral catalysts. Because of its great age, this oil can potentially yield valuable information about the size and diversity of the early biosphere, particularly if it contains molecular fossils (biomarkers) of the primordial organisms from which it was derived.

Bituminous nodules are common in several Archaean sandstones

from the Pilbara craton of northwestern Australia and the Kaapvaal craton of southern Africa⁶, as well as in the early Palaeoproterozoic Huronian Supergroup of the Superior craton, Canada⁷. The nodules occur in laminae rich in detrital heavy minerals and occupy intergranular pore space. As they contain remnants of radioactive detrital grains, they evidently formed in the same way as identical Phanerozoic nodules⁸, by radiogenic polymerization and cross-linking of fluid hydrocarbons in petroleum⁹. During metamorphism, pore space in the host sandstones was occluded and potential petroleum source rocks became overmature⁶. Hence, fluid petroleum precursors to the solid bituminous nodules must have been generated and immobilized before peak metamorphism, a mid-Palaeoproterozoic (or older) event in all cases.

To determine if any original petroleum fluids were still extant, we further investigated the best preserved of these bituminous sandstones. The samples, all of mature quartz sandstone, were obtained from drill cores or underground mines to avoid oxidation by surface weathering and to minimize contamination by modern organisms. They range in age from ~2,450 Myr (Matinenda Formation) to ~3,000 Myr (Lalla Rookh Formation) old (Table 1) and for most, metamorphism never exceeded prehnite–pumpellyite facies (200–300 °C). Though deposited in differing environments, each sandstone was intimately associated with organic-rich mudrocks that would have been hydrocarbon sources before metamorphic overmaturation. For samples containing bituminous nodules, double-polished thick sections were examined by transmitted-light (TL) and ultraviolet-epifluorescence (UV) microscopy (330–380 nm), the best non-destructive methods for identifying oil in fluid inclusions¹⁰.

Fluorescent fluid inclusions, all very small (<5–10 µm), are most common in the Canadian sample, fairly abundant in Kaapvaal samples, but are quite rare in Pilbara rocks where only about ten were found, perhaps because their faint fluorescence rapidly fades. The Canadian sample contains the greatest diversity of fluorescing inclusions, including the largest individuals. The most complex

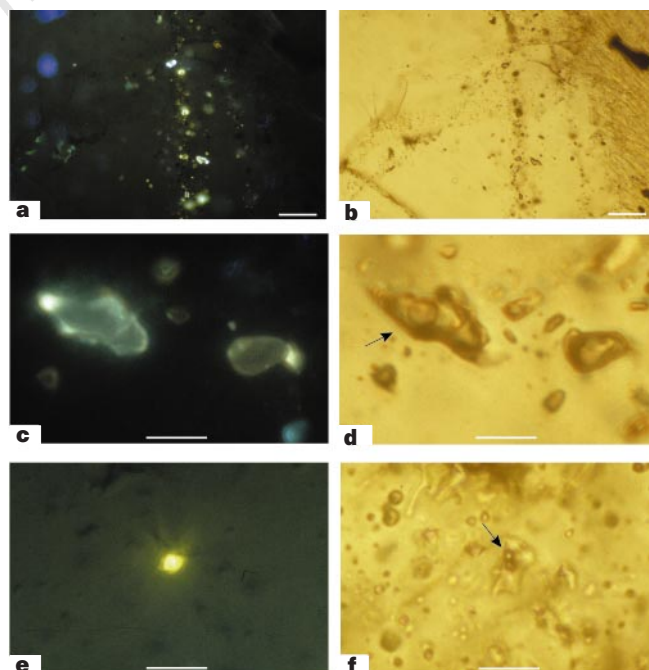


Figure 1 Photomicrographs showing range of fluid inclusion types and fluorescence colours in healed microfractures in Archaean and earliest Palaeoproterozoic sandstones. **a, b**, Trail of four-phase, $H_2O(l)$ -oil- $CO_2(l)$ - $CO_2(v)$ inclusions fluorescing white, blue-white and yellow (**a**, UV; **b**, TL), Matinenda Formation, Huronian Supergroup. **c, d**, Four-phase inclusions fluorescing blue-white and yellow comprising 20% $H_2O(l)$ (non-fluorescing outer rim), 15% oil

(fluorescent part) and 65% $CO_2(l)$ + $CO_2(v)$ (non-fluorescing inner portion) at 22 °C (**c**, UV; **d**, TL, gas bubble marked by arrow), Matinenda Formation, Huronian Supergroup. **e, f**, Orange-fluorescing, three-phase, water-oil-gas inclusion with oil forming a fluorescent rim around the gas bubble (**e**, UV; **f**, TL, gas bubble marked by arrow), Composite Reef, Elsburg Formation, Witwatersrand Supergroup. Scale bar: 100 µm in **a, b**, 10 µm in **c-f**.

Table 1 Location, geological setting and age of samples

Craton	Location	Group or supergroup	Formation or reef	Age (Myr)	Metamorphic age (Myr)
Pilbara	Gorge Creek	De Grey	Lalla Rookh	~3,000	3,000–2,750
Kaapvaal	Evander	Witwatersrand	Kimberley Reef	~2,850	2,550–2,050
Kaapvaal	Randfontein	Witwatersrand	Composite Reef	~2,850	2,550–2,050
Pilbara	Bronzewing Pool	Fortescue	Hardey	~2,760	2,450–2,100
Kaapvaal	Randfontein	Transvaal	Black Reef	~2,590	2,350–2,050
Superior	Elliot Lake	Huronian	Matinenda	~2,450	2,200–1,850

Data from refs 6, 7, 17, 32.

have four fluid phases (Fig. 1a–d), with a non-fluorescent liquid rim surrounding a fluorescent liquid which hosts a non-fluorescent liquid bubble containing a mobile gas bubble. The inner liquid and gas bubble homogenize to liquid at 26–29 °C, freeze between –114 and –101 °C and melt at –65 to –57 °C, consistent with CO₂. The outer non-fluorescing liquid freezes at –37 to –45 °C and melts about 0 °C, indicating water. The fluorescent liquid is oil, confirmed by Fourier-transform infrared spectroscopy (FTIR) of the largest inclusions which indicates the presence of aliphatic hydrocarbons, with fluorescence mainly induced by aromatic molecules¹¹. Coexistence of CO₂, oil and water in fluid inclusions, apparently not previously recorded, may be restricted to metamorphosed rocks. The Kaapvaal samples also contain a diverse suite of inclusions, including flat, elliptical and irregular individuals fluorescing blue, white or orange with two, three or four phases (Fig. 1e, f and Fig. 2a,

d). Pilbara inclusions, including those in the oldest rocks studied here, are mostly elliptical with two fluid phases (Fig. 3a, b): a spherical gas bubble within a dull blue-fluorescing liquid (oil).

To determine the relative timing of fluid inclusion emplacement and thus petroleum migration, the samples were examined by backscattered-electron (BSE) and cathodoluminescence (CL) scanning-electron microscopy (SEM). This showed that occasionally, petroliferous fluid inclusions were trapped by syntaxial quartz overgrowths enclosing detrital quartz grains (Fig. 2d–f). Most, however, occur within detrital quartz grains in healed microfractures (Fig. 2a–c, Fig. 3c–f) filled by optically identical secondary quartz (Fig. 3c, d). The microfractures, 1–50 µm wide, have random orientations and terminate at grain boundaries, even where the intergranular material is quartz cement in crystallographic continuity with the host grain (Fig. 3c, d). Where metamorphic sericite

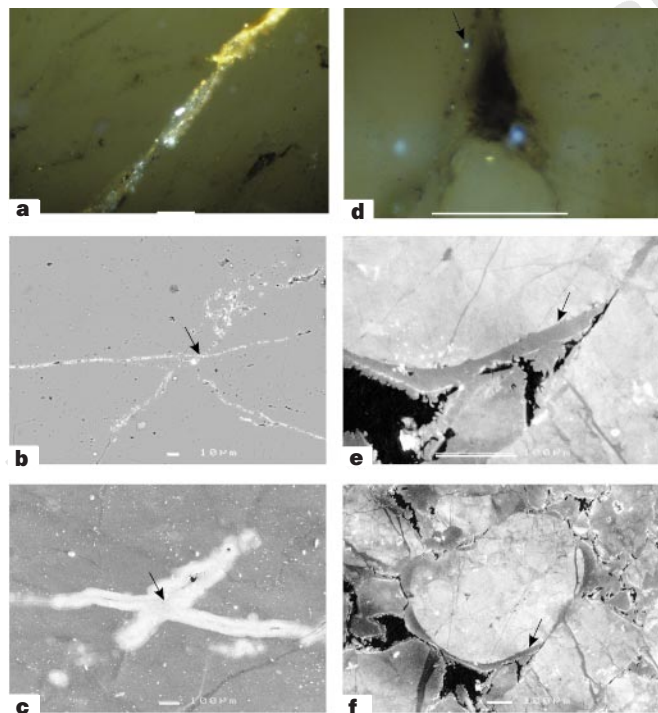


Figure 2 Photomicrographs and SEM images showing location of fluid inclusions in relation to grain boundaries, overgrowths and microfractures in Kaapvaal Archaean sandstones. **a–c**, White, orange and blue fluorescing fluid inclusions (**a**, UV) in microfractures also containing remobilized uraninite and pyrite (**b**, BSE, uraninite white, location of fluid inclusions marked by arrow) and surrounded by pale radiation-damaged detrital quartz (**c**, CL, location of fluid inclusions marked by arrow), Black Reef Formation, Transvaal Supergroup. **d–f**, Blue-fluorescing fluid inclusion (**d**, UV, fluid inclusion marked by arrow) on margin of detrital quartz grain (**e**, CL, location of fluid inclusion marked by arrow) overgrown by secondary quartz rim (**f**, CL, detrital quartz pale grey, secondary quartz mid-grey, interstitial sericite matrix black, location of fluid inclusion marked by arrow), Composite Reef, Elsburg Formation, Witwatersrand Supergroup. Scale bar, 10 µm in **a**, **b**, 100 µm in **c–f**.

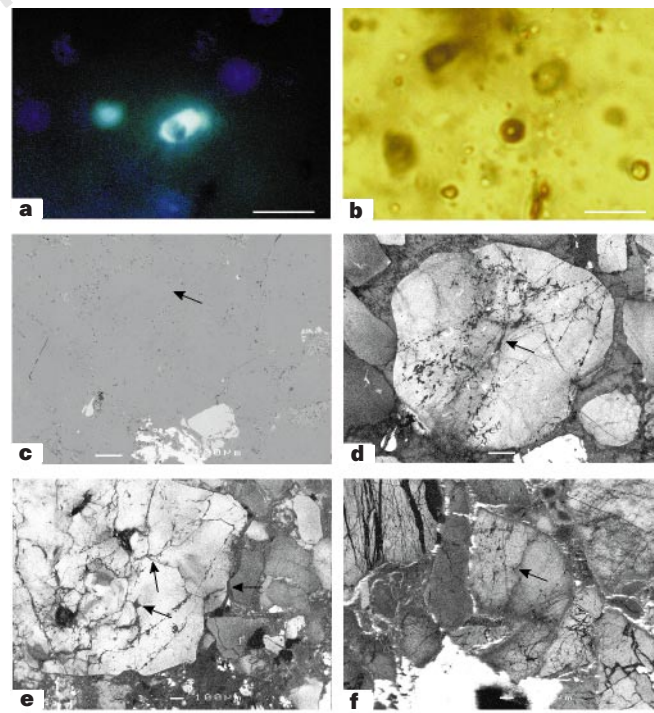


Figure 3 Photomicrographs and SEM images showing petrological setting of fluorescent fluid inclusions in Pilbara Archaean sandstones. **a**, **b**, Blue-fluorescing, two-phase, oil-gas inclusion (**a**, UV; **b**, TL), Lalla Rookh Formation, De Grey Group. **c**, **d**, Detrital quartz grain containing fluorescent fluid inclusions (locations marked by arrows) in microfractures terminating at grain margin (**c**, BSE showing that surrounding grains and cement are all quartz except for detrital potassic feldspar grain at bottom centre; **d**, CL showing non-luminescent quartz filling microfractures), Hardey Formation, Fortescue Group. **e**, **f**, CL images of detrital quartz grains containing fluorescent fluid inclusions (locations marked by arrows) in microfractures terminating at grain margins (**e**, note brecciation of surrounding grains—large arrow; **f**, note pressure solution of grain contacts—large arrow), Lalla Rookh Formation, De Grey Group. Scale bar, 10 µm in **a**, **b**, 100 µm in **c–f**.

surrounds detrital grains (Fig. 3e, f), it is not included within healed microfractures or quartz overgrowths, confirming that pore occlusion occurred before or during metamorphism.

The oil-bearing fluid inclusions enclosed by syntaxial quartz overgrowths were entrapped, at the latest, when pressure-solution provided abundant silica for cementation, an event that generally occurs at moderate burial depths (1–3 km) fairly early in diagenesis¹². Those in microfractures were probably emplaced even earlier, as such structures start forming during compaction at <1 km burial depth¹³ and heal soon after opening¹⁴. They require open pore-space for grain point-contacts to impinge and focus local stress, causing internal brittle failure, rotating framework grains and brecciating surrounding grains¹⁵. That the microfractures terminate at grain margins despite the continuous homogeneous medium now present (Fig. 3c, d) further indicates that pores were not completely cemented during fracturing¹⁶. So maximum burial and peak metamorphism, when primary pore space was obliterated⁶, represents the latest possible time for inclusion emplacement. In all instances, this occurred before 1,850 Myr ago (Table 1) and, in the Lalla Rookh Formation, during the Archaean aeon itself. For the Black Reef sample, the timing of oil entrapment can be estimated more accurately, because the host microfracture also contains secondary uraninite (Fig. 2d). Radiometric dating has shown that, in this formation, uranium minerals were remobilized along with hydrocarbons 2,380–2,330 Myr ago¹⁷.

The discovery of oil preserved in Archaean fluid inclusions extends the known geological range of liquid hydrocarbons well beyond the current limit of latest Palaeoproterozoic⁴. It confirms the contention⁶ that hydrocarbon generation by thermal maturation of biogenic kerogen was extensive in Archaean sedimentary basins. This is somewhat at odds with conventional wisdom about petroleum distribution, which regards Archaean terrains as essentially unpertoliferous. This view is based on the presumption that no suitably organic-rich source rocks occurred in the Archaean³, and that even if they did, subsequent metamorphism and deformation would have destroyed any petroleum they produced¹⁸. However, surveys of the organic carbon contents in Archaean sedimentary rocks¹⁹ clearly show that after accounting for metamorphism decarbonation, kerogen levels in many Archaean mudrocks were high enough to have yielded significant quantities of hydrocarbon on thermal maturation⁶. Moreover, the survival of oil for about 3 billion years, as documented here, shows that post-depositional trauma did not always destroy all traces of petroleum fluids.

The preservation of liquid hydrocarbons in Archaean rocks that have been regionally metamorphosed to prehnite–pumpellyite facies (200–300 °C) may seem paradoxical. Under normal circumstances, petroleum liquids quickly degrade to pyrobitumen and methane at temperatures around 200 °C (ref. 20). However, these limits may only apply to open systems under low pressures where circulating fluids and catalytic substrates can interact with the petroleum, such as in oil reservoirs or pore networks. In high-pressure closed systems with inert hosts, the kinetics of hydrocarbon degradation are sufficiently slow to allow oil to survive higher temperatures for considerable periods²¹. For instance, long-chain hydrocarbons have been conserved in fluid inclusions within Devonian hydrothermal vein quartz where temperatures temporarily reached 300–450 °C (ref. 22). That fluid inclusions can shelter complex hydrocarbons for about a billion years under cool cratonic thermal regimes is already well established^{23,24}. So the existence of oil in metamorphosed rocks billions of years old, as described here, provides strong evidence that time should no longer be regarded as a necessary parameter controlling petroleum degradation, and that the upper temperature limit for oil stability is rather higher than commonly presumed.

As hydrocarbon generation and migration were evidently widespread phenomena in Archaean and early Palaeoproterozoic basins, global burial of organic carbon must have been substantial during

these early stages of the Earth's history. Indeed, isotopic mass balances indicate that at least 12% of all buried Archaean carbon was organic²⁵. The amount of organic carbon buried depends on a combination of productivity and preservation. Although oxygen-poor subaqueous environments probably prevailed during the Archaean²⁶ and aerobic remineralization of organic matter would have thus been curtailed, there is abundant evidence for vigorous microbial recycling of sedimentary organic detritus by other metabolic pathways back to the late Archaean²⁷. Hence, if Archaean productivity were very low, microbial decomposition of sedimentary organic matter would probably have precluded kerogen levels high enough for successful expulsion of hydrocarbons from source rocks. Indeed, as nutrient levels are unlikely to have been strongly limiting in Archaean water bodies, with exposed continents²⁸ supplying phosphate and nitrogen-fixing cyanobacteria probably present²⁹, the Archaean aquatic biota may well have been as productive as its modern counterpart⁵. So despite marked growth of the crustal organic carbon reservoir during the Proterozoic³⁰, the hydrocarbon record implies that soon after life appeared on Earth it proliferated in aquatic environments. The diversity of this primordial biosphere may also be constrained by the oil in the fluid inclusions, should it prove to contain molecular fossils of the organisms from which it was derived (biomarkers³¹).

Received 22 January; accepted 28 July 1998.

1. Quigley, T. M. & Mackenzie, A. S. The temperatures of oil and gas formation in the subsurface. *Nature* **333**, 549–553 (1988).
2. Mango, F. D. The stability of hydrocarbons under the time-temperature conditions of petroleum genesis. *Nature* **352**, 146–148 (1991).
3. Bois, C., Bouche, P. & Pelet, R. Global geologic history and distribution of hydrocarbon reserves. *Bull. Am. Assoc. Petrol. Geol.* **66**, 1248–1270 (1982).
4. Jackson, M. J., Powell, T. G., Summons, R. E. & Sweet, I. P. Hydrocarbon shows and petroleum source rocks in sediments as old as 1.7×10^9 years. *Nature* **322**, 727–729 (1986).
5. McKirdy, D. M. & Imbus, S. W. in *Early Organic Evolution: Implications for Mineral and Energy Resources* (eds Schidlowski, M. et al.) 176–192 (Springer, Berlin, 1992).
6. Buick, R., Rasmussen, B. & Krapez, B. Archaean oil: evidence for extensive hydrocarbon generation and migration 2.5–3.6 Ga. *Bull. Am. Assoc. Petrol. Geol.* **82**, 50–69 (1998).
7. Mossman, D. J., Nagy, B. & Davis, D. W. Hydrothermal alteration of organic matter in uranium ores, Elliot Lake, Canada: implications for selected organic-rich deposits. *Geochim. Cosmochim. Acta* **57**, 3521–3529 (1993).
8. Rasmussen, B., Glover, J. E. & Alexander, R. Hydrocarbon rims on monazite in Permian-Triassic arenites, northern Perth Basin, Western Australia: pointers to the former presence of oil. *Geology* **17**, 115–118 (1989).
9. Rasmussen, B. Fluorescent growth bands in irradiated-bitumen nodules: evidence of episodic hydrocarbon migration. *Bull. Am. Assoc. Petrol. Geol.* **81**, 17–25 (1997).
10. Roedder, E. *Fluid Inclusions* (Mineral. Soc. Am., Washington DC, 1984).
11. Pironon, J. & Pradier, B. Ultraviolet-fluorescence alteration of hydrocarbon fluid inclusions. *Org. Geochem.* **18**, 501–509 (1992).
12. Bjørlykke, K. & Egeberg, P. K. Quartz cementation in sedimentary basins. *Bull. Am. Assoc. Petrol. Geol.* **77**, 1539–1548 (1993).
13. Zhang, J., Wong, T.-F. & Davis, D. M. Micromechanics of pressure-induced grain crushing in porous rocks. *J. Geophys. Res.* **95**, 341–352 (1990).
14. Brantley, S. L., Evans, B., Hickman, S. L. & Crerar, D. A. Healing of microcracks in quartz: implications for fluid flow. *Geology* **18**, 136–139 (1990).
15. Laubach, S. E. A method to detect natural fracture strike in sandstones. *Bull. Am. Assoc. Petrol. Geol.* **81**, 604–623 (1997).
16. Groshong, R. H. Low-temperature deformation mechanisms and their interpretation. *Geol. Soc. Am. Bull.* **100**, 1329–1360 (1988).
17. Robb, L. J., Charlesworth, E. G., Drennan, G. R., Gibson, R. L. & Tongu, E. L. Tectono-metamorphic setting and paragenetic sequence of Au-U mineralisation in the Archaean Witwatersrand Basin, South Africa. *Aust. J. Earth Sci.* **44**, 353–371 (1997).
18. Hobson, G. D. & Tiratsoo, E. N. *Introduction to Petroleum Geology* (Gulf Publ. Co., Houston, 1981).
19. Strauss, H. & Moore, T. B. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 709–798 (Cambridge Univ. Press, 1992).
20. Ungerer, P. State of the art of research in kinetic modelling of oil formation and expulsion. *Org. Geochem.* **16**, 1–25 (1990).
21. Price, L. C. Thermal stability of hydrocarbons in nature: limits, evidence, characteristics, and possible controls. *Geochim. Cosmochim. Acta* **57**, 3261–3280 (1993).
22. Hoffmann, C. F., Henley, R. W., Higgins, N. C., Solomon, M. & Summons, R. E. Biogenic hydrocarbons in fluid inclusions from the Aberfoyle tin-tungsten deposit, Tasmania, Australia. *Chem. Geol.* **70**, 287–299 (1988).
23. Newell, K. D., Burruss, R. C. & Palacas, J. G. Thermal maturation and organic richness of potential petroleum source rocks in Proterozoic Rice Formation, North American Mid-Continent rift system, northeastern Kansas. *Bull. Am. Assoc. Petrol. Geol.* **77**, 1922–1941 (1993).
24. George, S. C. & Jardine, D. R. Ketones in a Proterozoic dolerite sill. *Org. Geochem.* **21**, 829–839 (1994).
25. Hayes, J. M. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 291–301 (Princeton Univ. Press, 1983).
26. Knoll, A. H. & Holland, H. D. in *Effects of Past Global Change on Life* (ed. Stanley, S.) 1–17 (National Academy of Science Press, Washington DC, 1995).
27. Hayes, J. M. in *Early Life on Earth: Nobel Symposium No. 84* (ed. Bengtson, S.) 220–236 (Columbia Univ. Press, New York, 1994).
28. Buick, R. et al. Record of emergent continental crust ~3.5 billion years ago in the Pilbara Craton of Australia. *Nature* **375**, 574–577 (1995).
29. Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of

- CO₂ in the ocean. *Nature* **387**, 272–275 (1997).
30. Des Marais, D. J., Strauss, H., Summons, R. E. & Hayes, J. M. Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment. *Nature* **359**, 605–609 (1992).
31. Summons, R. E. & Walter, M. R. Molecular fossils and microfossils of prokaryotes and protists from Proterozoic sediments. *Am. J. Sci.* **290-A**, 212–244 (1990).
32. Martin, D. McB., Clendenin, C. W., Krapez, B. & McNaughton, N. J. Tectonic and geochronological constraints on late Archaean and Palaeoproterozoic stratigraphic correlation within and between the Kaapvaal and Pilbara Cratons. *J. Geol. Soc. Lond.* **155**, 311–322 (1998).

Acknowledgements. We thank G. Ellis, T. S. Blake and B. Krapez for samples, M. Kassis for sample preparation, T. Vassallo for FTIR interpretation and P. J. Eadington, R. D. Müller, L. C. Ivany and B. Krapez for discussions and criticism. This work was supported by an ARC grant to R.B. and an ARC postdoctoral fellowship to B.R.

Correspondence and requests for materials should be addressed to R.B. (e-mail: buick@es.us.oz.au).

Triggering of volcanic eruptions

Alan T. Linde & I. Selwyn Sacks

Department of Terrestrial Magnetism, Carnegie Institution of Washington,
5241 Broad Branch Road NW, Washington DC 20015, USA

Although earthquakes and volcanic eruptions are each manifestations of large-scale tectonic plate and mantle motions, it is usually thought that the occurrences of these events are not directly related. There have been some studies, however, in which triggering of volcanic eruptions by earthquakes (remote from the volcano) has been proposed^{1,2}. The 1992 Landers (southern California) earthquake caused triggered seismicity at very large distances³, including the magmatically active⁴ Long Valley caldera region which also experienced a significant coincident deformation transient⁵. Motivated by this demonstration of the ability of a distant earthquake to disturb a volcanic system, and the earlier studies of specific cases of eruption triggering, we examine here the historical record of eruptions and earthquakes to see if there are indeed significantly more eruptions immediately following large earthquakes. We find that within a day or two of large earthquakes there are many more eruptions within a range of 750 km than would otherwise be expected. Additionally, it is well known⁶ that volcanoes separated by hundreds of kilometres frequently erupt in unison; the characteristics of such eruption pairs are also consistent with the hypothesis that the second eruption is triggered by earthquakes associated with the first.

Small-to-moderate (magnitudes up to ~6) earthquakes usually accompany volcanic eruptions; they may occur before and after the initial eruption as magma movement stresses the surrounding crust. Our concern here is rather with large earthquakes distant from the volcano, and with the activity associated with other eruptions. We are aware of only a few studies in which triggering of volcanic eruptions by earthquakes is considered, even though the possibility had been raised long ago⁷. Yokoyama¹ and Nakamura² have collected many reports of volcanic activity (including volcanic earthquakes, explosions, steamings and rumblings) which strongly suggest a connection and Nakamura gives a table of proposed triggered eruptions. They proposed that either the static strains or the seismic waves may have caused the triggering. It has been suggested that eruptions in Java⁸ and South America⁹ are triggered by earthquakes. Williams¹⁰ commented on activity at Ulawun within a week of the eruption and earthquake at Rabaul (120 km away) in 1994 and suggested that a bubble mechanism^{11,12}, as proposed by Linde *et al.*⁵, could apply. Gudmundsson and Saemundsson¹³ reported on a weak relationship between earthquakes and eruptions in Iceland. Caldera unrest¹⁴ often follows large earthquakes. Triggered seismicity, at distances in excess of 1,000 km, was dramatically clear following the Landers earthquake³. Also clear from borehole strainmeter and long-baseline tiltmeter records⁵ is that the Long Valley caldera experienced a significant deformation

transient coherent with the enhanced rate of local seismicity. Although it has been noted⁶ that separated volcanoes frequently erupt on the same day, we are unaware of any mechanism proposed to explain this.

Here we take an approach complementary to that of searching for direct evidence^{1,2} for the pairing of earthquakes and eruptions. We examine global catalogues for the past several hundred years. We use the Smithsonian Institution, Global Volcanism Program catalogue⁶ of volcanic eruptions, and combine the earthquake catalogue from the US Geological Survey National Earthquake Information Center compendium¹⁵ (duplication of events is avoided). The Landers experience indicates that the process being considered is more likely to be activated by large earthquakes—previous smaller earthquakes in California did not generate such clear triggering. Thus we consider great (magnitude ≥ 8) earthquakes and also those with magnitudes in the range 7.0–7.9. We also require that the eruption be within a few days of the earthquake in order to be a candidate triggered eruption; earlier investigations have accepted longer time windows. From the eruption catalogue, we consider only those for which the date is given to the day; many eruptions occur in remote areas and the start time of the eruption is often not known. We also limit the search to eruptions with volcanic explosivity index (VEI, a 0–8 scale of explosive magnitude) of 2 or greater; this excludes non-explosive and small eruptions and so adds to the robustness of our

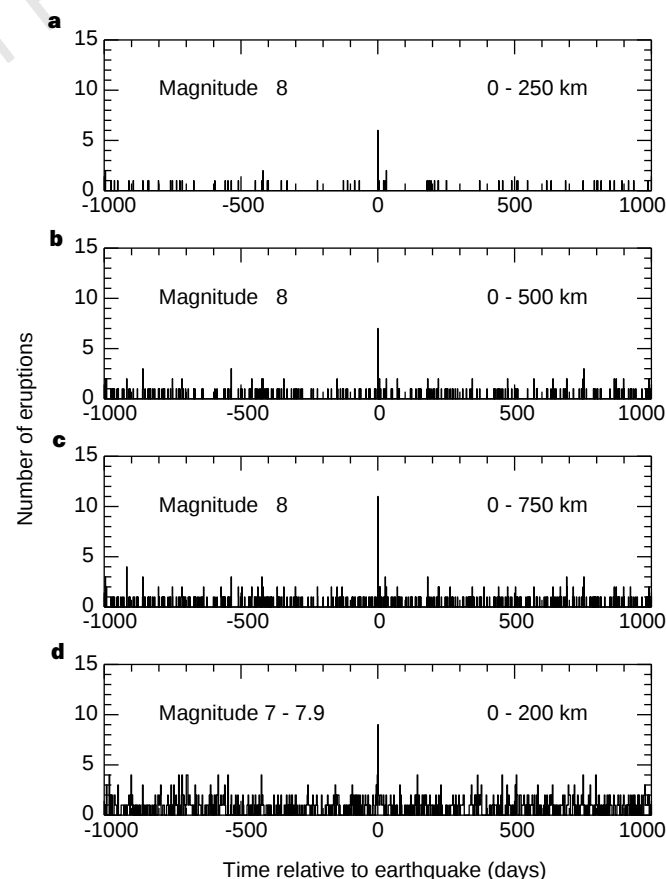


Figure 1 Histograms of number of eruptions occurring within $\pm 1,000$ days of large earthquakes. In **a**, **b** and **c**, the eruptions, following earthquakes of magnitude ≥ 8 , are grouped in 1-day bins; **a**, **b** and **c** are for earthquake–eruption distances of up to 250 km, 500 km and 750 km, respectively. The large peak for the bin following the earthquake is evidence for a triggering effect, although there is no evidence for triggering in the distance range 250–500 km. In **d**, the eruptions following earthquakes of magnitude 7–7.9 are grouped in 2-day bins for the distance range 0–200 km. Most of these eruptions are at distances > 100 km. All other time bins display expected random values.

Table 1 List of earthquake-triggered eruption pairs in Fig. 1

Earthquake							Dist. (km)	Eruption							
Year	Local time			Lat. (deg.)	Long (deg.)	Mag.		Type*	Year	Time		VEI†	Lat. (deg.)	Long (deg.)	Volcano‡
	Month	Day	Hour							Month	Day				
1587	09	03	20	−0.22	−78.50	7.7	12	P	1587	09	03	3	−0.17	−78.60	Guargua Pichincha (Ecuador)
1730	07	08	05	−33.05	−71.63	8.7	707	S	1730	07	08	2	−39.42	−71.93	Villarrica (Chile-C)
1737	12	24	00	−39.80	−73.20	7.7	117	S	1737	12	24	2	−39.42	−71.93	Villarrica (Chile-C)
1822	11	19	21	−33.05	−71.63	8.5	180	S	1822	11	19	2	−33.78	−69.90	San Jose (Chile-C)
1822	11	19	21	−33.05	−71.63	8.5	707	S	1822	11	19	2	−39.42	−71.93	Villarrica (Chile-C)
1835	02	20	10	−36.83	−73.03	8.1	663	S	1835	02	20	2	−42.78	−72.43	Minchinmavida (Chile-S)
1835	02	20	10	−36.83	−73.03	8.1	732	S	1835	02	20	2	−43.42	−72.83	Yanteles, Cerro (Chile-S)
1837	11	07	06	−39.80	−73.20	8.0	117	S	1837	11	07	2	−39.42	−71.93	Villarrica (Chile-C)
1837	11	07	06	−39.80	−73.20	8.0	156	S	1837	11	07	2	−41.10	−72.49	Osorno (Chile-S)
1854	06	27	00	50.50	158.00	7.0	178	S	1854	06	27	2	50.86	155.55	Alaid (Kurile Is.)
1868	08	15	15	0.80	−77.70	7.0	183	S	1868	08	15	2	−0.68	−78.44	Cotopaxi (Ecuador)
1899	09	03	20	60.00	−142.00	8.3	249	S	1899	09	03	2	62.00	−144.02	Wrangell (Alaska-E)
1913	03	14	16	4.50	126.50	8.3	144	S	1913	03	14	2	3.67	125.50	Awu (Sangihe Is. –Indonesia)
1913	10	14	19	−19.50	169.00	8.1	371	S	1913	10	14	2	−16.25	168.12	Ambrym (Vanuatu)
1914	01	12	18	31.50	131.00	7.0	153	S	1914	01	13	2	32.88	131.10	Aso (Kyushu –Japan)
1920	12	09	23	−39.00	−73.00	7.4	104	S	1920	12	10	2	−39.42	−71.93	Villarrica (Chile-C)
1934	12	31	10	32.00	−114.75	7.1	121	S	1934	12	31	2	31.77	−113.50	Pinacate Peaks (Mexico)
1939	01	30	12	−6.50	155.50	7.9	52	S	1939	01	30?	2	−6.14	155.20	Bagana (Bougainville Is.)
1939	04	30	12	−10.50	158.50	8.1	174	S	1939	04	30?	2	−9.02	157.95	Kavachi (Solomon Is.)
1974	01	10	19	−14.43	166.86	7.2	71	P	1974	01	11	2	−14.27	167.50	Gaua (Vanuatu)

* Type of eruption is either S for start or P for paroxysmal event.

† VEI is a measure of the explosive magnitude of the eruption.

‡ In volcano region, C denotes central; S denotes south; E denotes east.

results. (Similar, but more pronounced, results obtain if we include all eruptions.) Our procedure is simple: for each earthquake we search the eruption catalogue for events within various distance and time separations. We stack the results for all earthquakes, so the final result of the search process is a matrix of values of number of eruptions as a function of time and distance relative to the earthquakes.

The results of this processing are shown in Fig. 1; Fig. 1a–c is for eruptions occurring within 250 km, 500 km and 750 km, respectively, of great earthquakes. For all of these ranges, there is a peak in the number of eruptions on the same day as the earthquake, although there is no incremental increase in the 250–500 km range. The ratio of this peak amplitude to the average in the other bins (the background rate) is very high. There is no enhanced eruption rate at distances greater than 750 km. Division of the distance intervals into smaller increments shows that these peaks are derived dominantly from the 100–250 km and the 600–750 km ranges, with no contribution from distances less than 117 km.

The relatively large number of eruptions that occur soon after a great earthquake appear to be a clear indication that some of these earthquakes have triggered eruptions. Our catalogue contained 204 great earthquakes, eight of which have triggered eruptions.

We have repeated this search process for earthquakes with magnitude 7–7.9 (Fig. 1d). In this case, the peak eruption count for the 2 days following the earthquakes is evident for distances as large as 200 km (9 eruptions). This observation of a larger number of eruptions within 2 days of the earthquakes is consistent with the fact that at Long Valley the deformation excursion took several days to reach its maximum. Table 1 is a list of all earthquake–eruption pairs in the peak groups in Fig. 1.

One difficulty in grouping these eruptions in time bins arises because eruption times are reported only to the day, and thus there can be some ambiguity as to which event was first. Figure 1 shows the counts assuming that all same-day eruptions follow the earthquake. In some cases (March 1913, T. Simkin, personal communication; both 1835, Darwin¹⁶) it is known that the eruption followed the earthquake; we know of no cases to the contrary in our list. Even if we were to assume that those same-day events for which we have no definite discriminating information are divided equally between before and after the earthquake, then the peaks in the histograms would be many times the background rate on the day following and would still be significant. We would also have a

smaller, but significant, peak for the day preceding the earthquake. As there is other work supporting the triggering of eruptions, but we are unaware of evidence for the converse, we assign the same-day eruptions as following, but our conclusions do not hinge on this assignment. Additionally, there may be concern that a following eruption is noted only because of heightened awareness after the earthquake shaking. From the database it is impossible to determine whether this is significant, but where we do have information it is clear that the eruptive activity began after the earthquake. Also arguing against a spurious effect is that most triggered eruptions are more than 100 km from trigger earthquakes with magnitude <8 and many are more than 500 km from events of magnitude ≥8.

The most direct, assumption-free method for determining whether such a high bin count could occur by chance is to use a variation of the bootstrap method of computer intensive analysis¹⁷. We use the actual eruption catalogue and a randomized earthquake catalogue. Because the catalogue does not have a uniform temporal distribution of events (many more big earthquakes listed for this century because of the dramatic change in detection and location capabilities), we change the origin time of each earthquake by a time

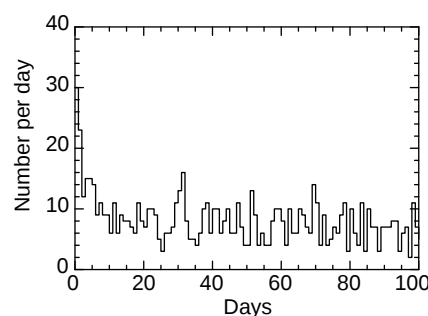


Figure 2 Histogram showing number of eruptions per day following a previous eruption and within 200 km of the first eruption. This distribution has a mean of 8.15 eruptions per day, with a standard deviation of 4.0. Paired eruptions, within two days, are more common than the average rate; the number of eruptions on the same day and following days are 5.4σ and 3.7σ above average. No such enhancement is observed for separations >200 km, and the effect decreases with separation up to 200 km.

increment that varies randomly in the range ± 100 days. This ensures that our random samples have the same overall distribution of events with time. Actual locations are retained. Using exactly the same procedure as for the real data, we ran 100,000 trials, with a $\pm 1,000$ -day time span, and determined the largest count in the 2,000 bins for each run. The largest count generated in any time bin was 8 (once), so that the probability of getting, by chance, a bin count of 8 with the real data is 10^{-5} ; the data set has a bin count of 11 for the day following the earthquakes. If we consider the event distributions to be poissonian, we reach comparable conclusions.

We have also examined the eruption catalogue for paired eruptions. Figure 2 shows that volcanoes separated by less than 200 km have paired (within two days) eruptions with frequencies much higher than for the background rate. As volcanic earthquakes are usually not large (infrequently as big as magnitude 6), these time–distance characteristics for the enhanced paired eruptions are consistent with our hypothesis that the second eruption is triggered by the passage of seismic waves. (For the time period examined, the earthquake catalogue is not complete, especially for smaller magnitudes, so we cannot identify the triggering earthquakes.)

The interval between eruptions for many volcanoes is usually decades or longer. If the recharge rate of the shallow reservoir is essentially constant with time, then the volcano can be at a near critical level for years. As the Landers earthquake produced an increase in pressure in the magma reservoir under Long Valley at a distance of about 400 km, then the same should be true generally: seismic waves from earthquakes have the potential to increase the pressure in magma chambers even at large distances from large earthquakes. For a volcano already close to the critical pressure state, this could result in a premature eruption. This conclusion is based on secure observations and does not depend on which (if any) of the various models^{5,18–23} proposed is correct, although we note that our suggestion⁵ is the only one that leads naturally to the timescale and character of the observed deformation.

The work emphasizes the need for a programme of continuous deformation monitoring of active volcanoes on a global scale, with sufficient sensitivity to be able to detect deformation transients on these short timescales. We note that a two-colour laser ranging system, with resolution of about 1 mm, was not able to detect changes during the period of triggered seismicity at Long Valley; seismic monitoring is important but gives only indirect information about the deformational changes. Very few active volcanoes are currently monitored well enough to allow detection of potentially important transient deformations. □

Received 14 April; accepted 14 July 1998.

- Yokoyama, I. Volcanic eruptions triggered by tectonic earthquakes. *Geophys. Bull. Hokkaido Univ.* **25**, 129–139 (1971).
- Nakamura, K. Volcano structure and possible mechanical correlation between volcanic eruptions and earthquakes. *Bull. Volcanol. Soc. Jpn* **20**, 229–240 (1975).
- Hill, D. P. *et al.* Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **260**, 1617–1623 (1993).
- Langbein, J., Hill, D. P., Parker, T. N. & Wilkenson, S. K. An episode of reinflation of the Long Valley Caldera, eastern California; 1989–1991. *J. Geophys. Res.* **98**, 15851–15870 (1993).
- Linde, A. T., Sacks, I. S., Johnston, M. J. S., Hill, D. P. & Bilham, R. G. Increased pressure from rising bubbles as a mechanism for remotely triggered seismicity. *Nature* **371**, 408–410 (1994).
- Simkin, T. & Siebert, L. *Volcanoes of the World* (Geoscience, Tucson, 1994).
- Michell, J. Conjectures concerning the cause, and observations upon the phenomena of earthquakes; particularly of that great earthquake of Nov. 1, 1755, which proved so fatal to the city of Lisbon, and whose effects were felt as far as Africa, and more or less throughout almost all of Europe. *Phil. Trans. R. Soc. Lond.* **51**, 566–634 (1760).
- Gulyas, E., Deak, G. & Hedervari, P. in *Proc. Symp. on Andean and Antarctic Volcanology Problems* (ed. Gonzalez, F. O.) 562–576 (Int. Assoc. Volcanol. and Chem. Earth's Interior, Rome, 1976).
- Gonzalez, F. O. Volcanic eruption related with great earthquakes as a parameter for forecasting event; the historical cases along the Chilean Andes. *Int. Un. Geodesy and Geophysics, General Assembly Abstr.* Vol. 2, 428 (Vancouver, 1987).
- Williams, S. N. Erupting neighbors; at last. *Science* **267**, 340–341 (1995).
- Steinberg, G., Steinberg, A. & Merzhanov, A. Fluid mechanism of pressure growth in volcanic (magmatic) systems. *Mod. Geol.* **13**, 257–265 (1989).
- Sahagian, D. L. & Proussevitch, A. A. Bubbles in volcanic systems. *Nature* **359**, 485 (1992).
- Gudmundsson, G. & Saemundsson, K. Statistical analysis of damaging earthquakes and volcanic eruptions in Iceland from 1550–1978. *J. Geophys.* **47**, 99–109 (1980).
- Newhall, C. G. & Dzurisin, D. Historical unrest at large calderas of the world. *US Geol. Surv. Bull.* **1855**, 19–20 (1988).
- Global hypocenter data base CD-ROM Version 3.0 (National Earthquake Information Center, US Geological Survey, Denver, 1994).

- Darwin, C. On the connexion of certain volcanic phenomena in South America; and on the formation of mountain chains and volcanoes, as the effect of the same power by which continents are elevated. *Trans. Geol. Soc. Lond.* **5**, (2nd ser.), 601–631 (1840).
- Diaconis, P. & Efron, B. Computer-intensive methods in statistics. *Sci. Am.* **248**, 116–130 (1983).
- Bodin, P. & Comberg, J. Triggered seismicity and deformation between the Landers, California, and Little Skull Mountain, Nevada, earthquakes. *Bull. Seismol. Soc. Am.* **84**, 835–843 (1994).
- Gomberg, J. & Bodin, P. Triggering of the $M_s = 5.4$ Little Skull Mountain, Nevada, earthquake with dynamic strains. *Bull. Seismol. Soc. Am.* **84**, 844–853 (1994).
- Anderson, J. G. *et al.* Seismicity in the western Great Basin apparently triggered by the Landers, California, Earthquake, 28 June 1992. *Bull. Seismol. Soc. Am.* **84**, 863–891 (1994).
- Hill, D. P., Johnston, M. J. S., Langbein, J. O. & Bilham, R. Response of Long Valley caldera to the $M_w = 7.3$ Landers, California, earthquake. *J. Geophys. Res.* **100**, 12985–13005 (1995).
- Johnston, M. J. S. *et al.* Transient deformation during triggered seismicity from the 28 June 1992 $M_w = 7.3$ Landers earthquake at Long Valley volcanic cinder cone, California. *Bull. Seismol. Soc. Am.* **85**, 787–795 (1995).
- Sturtevant, B., Kanamori, H. & Brodsky, E. E. Seismic triggering by rectified diffusion in geothermal systems. *J. Geophys. Res.* **101**, 25262–25282 (1996).

Acknowledgements. We thank T. Simkin for making available the Smithsonian eruption catalogue in digital form.

Correspondence and requests for materials should be addressed to A.T.L. (e-mail: linde@dtm.ciw.edu).

Electric fish measure distance in the dark

Gerhard von der Emde*, Stephan Schwarz*, Leonel Gomez†‡, Ruben Budelli‡ & Kirsty Grant†

* *Institute für Zoologie, Universität Bonn, Poppelsdorfer Schloss, D-53115 Bonn, Germany*

† *Institute Alfred Fessard, C.N.R.S., F-91198 Gif-sur-Yvette, Cedex, France*

‡ *Department of Biomathematics, Faculty of Science, University of the Republic of Uruguay, Montevideo, Uruguay*

Distance determination in animals can be achieved by visual or non-visual cues¹. Weakly electric fish use active electrolocation for orientation in the dark². By perceiving self-produced electric signals with epidermal electroreceptors, fish can detect, locate and analyse nearby objects. Distance discrimination, however, was thought to be hardly possible because it was assumed that confusing ambiguity could arise with objects of unknown sizes and materials^{3–5}. Here we show that during electrolocation electric fish can measure the distance of most objects accurately, independently of size, shape and material. Measurements of the 'electric image' projected onto the skin surface during electrolocation^{6–8} revealed only one parameter combination that was unambiguously related to object distance: the ratio between maximal image slope and maximal image amplitude. However, slope-to-amplitude ratios for spheres were always smaller than those for other objects. As predicted, these objects were erroneously judged by the fish to be further away than all other objects at an identical distance. Our results suggest a novel mechanism for depth perception that can be achieved with a single, stationary two-dimensional array of detectors.

For orientation in their environment, animals use sensory information to judge the distance of an object. Disparities in the two retinal images of an object can give stereoscopic depth perception in animals whose eyes have overlapping fields of vision^{9,10}. Other mechanisms for distance discrimination rely on the movement of detectors and successive comparison of images¹¹, on binocular convergence, or on measuring the speed at which objects' images move across the retina¹². Nocturnally active animals, however, must supplement vision with other senses such as hearing, mechanoreception or olfaction^{13–15}.

Weakly electric fish orient themselves at night by using active electrolocation^{2,16}. By emitting an electric signal (Fig. 1b) with a specialized electric organ, they generate an electric field and perceive the resulting current that flows through epidermal electroreceptors. Nearby objects distort the electric field and alter the current flowing through the electroreceptors⁶. Analysis of the 'electric image', that is, the spatial pattern of voltage change, that objects project onto the

skin surface^{7,8} enables a fish to detect and locate objects in complete darkness, and also to assess their electrical properties³. However, unequivocal distance perception was thought not to be possible without auxiliary means, because differences in the sizes and materials of objects could affect the electric image in the same way as differences in distance^{4,17}. In contrast to visual images, which become smaller with an increase in object distance, an increase in distance during active electrolocation leads to a larger size of the electric image¹⁸ and a decrease in its maximal amplitude (intensity) (Fig. 1a). However, a similar but larger object at the same distance would have the same effect. Thus, if fish only measured image size or image amplitude, they should not be able to determine the distances of unknown objects of different sizes.

Experiments were conducted with the weakly electric fish *Gnathonemus petersii*, a nocturnal, stream dwelling predator of insect larvae¹⁹, to find out whether, and how accurately, electric fish can measure object distance. Over a period of 3–5 weeks, three fish were trained to electrolocate two objects according to their distance behind corresponding fixed gates, and to swim through the gate behind which the object was further away (Fig. 1c). Fish alternated between the two gates several times before making a decision, thereby scanning each object with electroreceptor arrays on both the sides of the body and the head.

First, we offered the fish two identical objects, which differed only in their distances from the gate. The distance of the closer object remained fixed in any given trial, defining 'gate distance'. In total, we tested each fish with 12 different combinations of objects, each tested at a minimum of four different gate distances. Discrimination was always correctly achieved provided that inter-object distances differed by more than a critical threshold value (Fig. 2a, b). Discrimination thresholds were correlated with gate distance: fish were better at determining the distances of nearby objects. When gate distances were held constant, thresholds were inversely correlated with object size: distance estimates improved when larger objects were offered.

Because the discrimination of identical objects in different positions could be based on measurements of electric image amplitude or size rather than on object distance, in subsequent experiments we offered the fish two objects that differed in size, shape (for example, a cube and a pyramid) or material (for example, a metal cube and a plastic cube). Twenty-five combinations of different types of object were tested. Differences in these physical parameters did not affect distance discrimination performance (Fig. 2c). Fish could always determine which object was more distant irrespective of size, shape or material of the objects, provided that the distance difference was above the threshold.

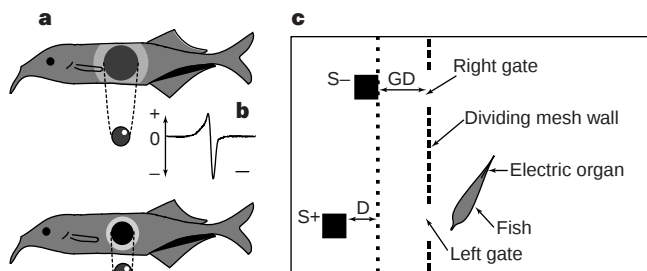


Figure 1 **a**, Electric image size and amplitude change with object distance. Electric images have a 'Mexican hat' profile⁷. An object far from the fish's skin (top) produces a relatively large electrical image with a small amplitude and low edge contrast, indicated by shading. When the same object comes closer (bottom), the size of the image decreases and maximal amplitude and edge contrast increase. **b**, Waveform of the electric organ discharge (EOD) of *Gnathonemus petersii*. Time scale bar equals 100 μ s. **c**, Set-up for behavioural experiments (not drawn to scale). GD, gate distance; D, distance between objects; S+, target object (farther away); S-, non-target object.

To understand which parameters of the electric image are used by fish to determine object distance, we measured the 'electric images' produced locally at the level of the electroreceptors by different objects (Fig. 3a, b). No single physical parameter showed a pattern of variation sufficient to explain distance discrimination⁵. Thus, if the fish had relied only on image size, amplitude or edge contrast, they would have made systematic errors when certain object combinations were offered. But such errors did not occur. Neither could fish have monitored the change of one of these parameters during their approach to the objects^{11,20}, because this would not have provided the fish with unequivocal distance cues when approaching an object of unknown size or impedance⁶. The only parameter combination that could provide a quantitative explanation compatible

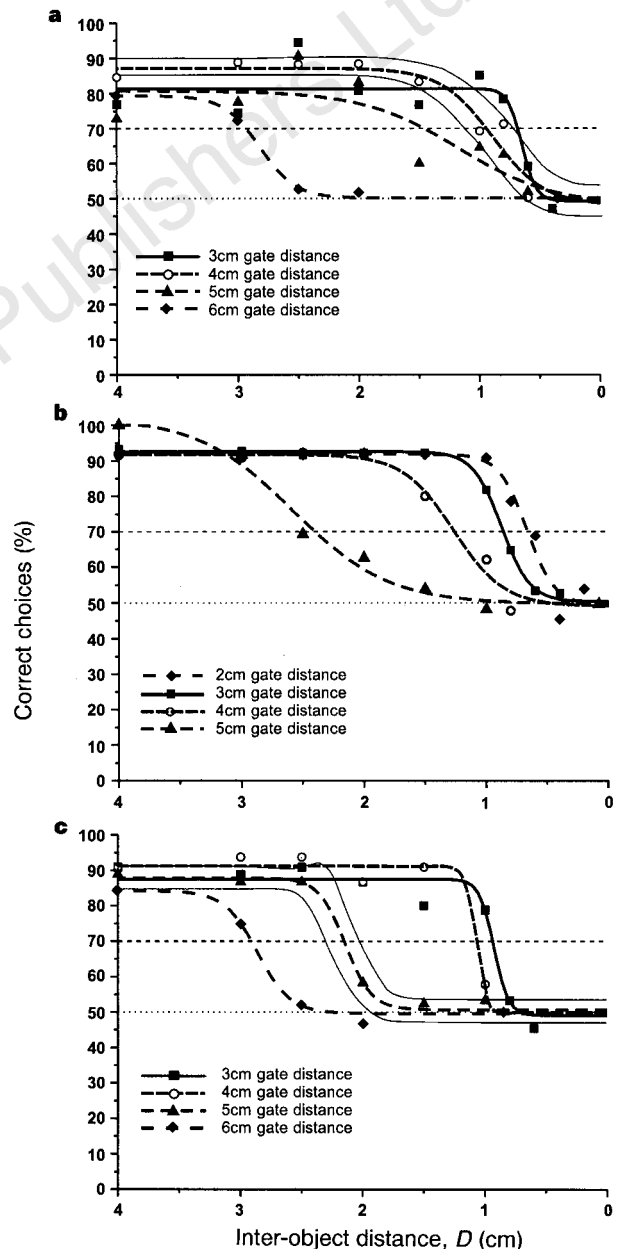


Figure 2 Psychometric functions for distance discrimination of **(a)** two identical metal cubes (64 cm³), **(b)** two identical metal spheres (33.5 cm³) and **(c)** two metal cubes of different sizes (125 and 27 cm³). Discrimination threshold was defined as the inter-object distance D for which the sigmoid function fitted to each data set crossed the 70% correct choice value. Each data point is based on at least 150 trials. In **(a)** and **(c)** thin lines indicate 95% confidence limits for two different curves (4 cm and 5 cm gate distances)³⁰. All curves are significantly different ($P < 0.05$).

with all our behavioural results was the ratio of the maximal slope of the 'electric image' to its maximal amplitude. This slope-to-amplitude ratio depended only on object distance and was independent of size, shape and material for most objects (Fig. 3c, d).

Our measurements also showed that only the rostral slope of the image (the left slopes in Fig. 3b, c) provided adequate information for the fish to determine object distance. The caudal (right) image slopes of different electric images varied unsystematically with object distance, probably because the proximity of the electric organ produces asymmetries in the electric field. Thus, they could not provide any reliable distance cues. We therefore suggest that fish used the rostral image slopes to determine slope-to-amplitude ratios and thus to determine object distance.

This hypothesis was tested in the following way. Electric image measurements showed that slope-to-amplitude ratios for spheres were smaller than those obtained for solid cubes of the same diameter at the same distance (Fig. 3c, d). The difference in slope-to-amplitude ratios corresponded to a difference in distance of ~ 1.5 cm at a gate distance of 3 cm (Fig. 3d, arrow). Thus, it was predicted that fish would erroneously judge a sphere to be ~ 1.5 cm further away than a cube at the same distance.

Experimentally we found this did in fact occur (Fig. 4). Three fish were given the task of comparing the distances of a metal sphere and a metal cube. When the sphere was further away, fish chose it correctly, but they also continued to choose it when both objects were at the same distance. At a 'gate distance' of 3 cm, when the cube was farther away, fish started to choose randomly between the two objects at an inter-object distance of ~ 2 cm. As inter-object distance was further reduced, fish incorrectly chose the sphere instead of the

cube (Fig. 4a). Thus, fish judged the sphere to be farther away than the cube, even though this was not true. The psychometric functions established at a 'gate distance' of 3 cm for distance discrimination of (1) a cube and a sphere and (2) two identical cubes (Fig. 4a) show that the curve for the cube/sphere discrimination crosses the 50% correct line ~ 1.5 cm closer to larger inter-object distances than that for the two cubes (Fig. 4a, two-headed arrow). Thus, at a 'gate distance' of 3 cm, the sphere seemed to the fish to be ~ 1.5 cm farther away than the cube as predicted from our electric image measurements (Fig. 3d). Similar 'electrical illusions' also occurred at the other 'gate distances' tested (2 and 4 cm; Fig. 4b), indicating a general mechanism for distance perception.

The results of our experiments suggest that electric fish can measure three-dimensional depth by analysing the electric image of objects projected onto a single, stationary two-dimensional array of electroreceptors. This mechanism differs from all spatial orientation mechanisms shown so far in other species, but resembles a mechanism hypothesized from theoretical postulates of hydrodynamical imaging by the lateral-line system of blind cave fish²¹. Other examples of stationary detector systems include auditory location of a sound source by the cod¹³, surface-feeding fish that determine distance by analysis of the frequency modulation and the curvature of a wave train²², and chameleons, toads and other species, which use the accommodation of their lenses to judge the distance of near objects^{10,23}. Echolocating bats, like electric fish, employ an active location system, but this is based on acoustics. In contrast to electric fish, bats can make time measurements²⁴ and thus can determine object distance by employing temporal cues^{25–27}. The mechanism proposed for distance measurement in electric fish

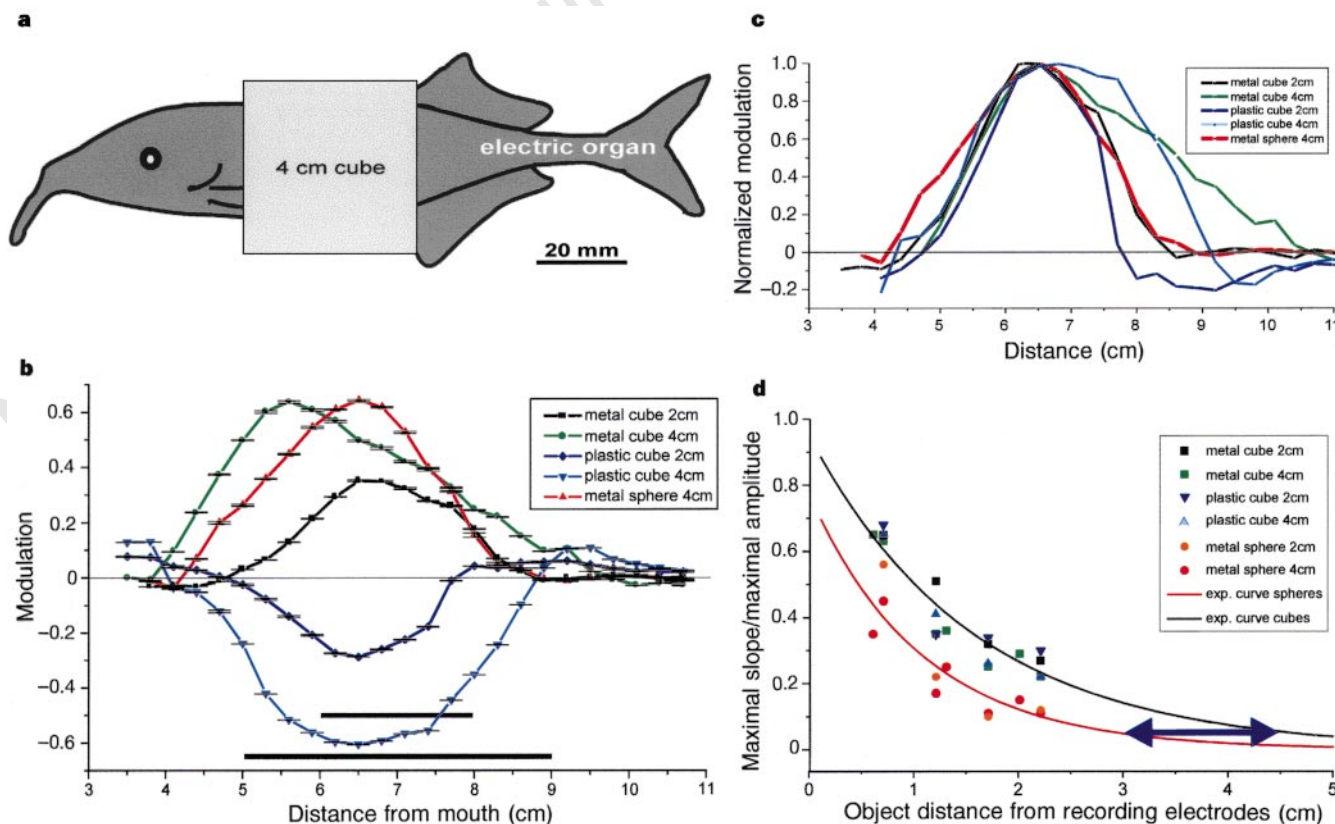


Figure 3 Electric image measurements. **a**, Fish with a cube positioned for electric image measurements. **b**, Electrical images projected by five different objects, 12 mm from the skin surface. Horizontal black bars mark the object position. **c**, Normalized electric images obtained from data shown in (**b**). Curves were shifted horizontally to align rostral (left) slopes approximately. Note that the rostral slopes for all cube-images are similar, whereas the slope for the sphere (red curve) is less steep. **d**, Slope-to-amplitude ratios calculated for two metal spheres (red curve)

and four cubes (black curve) plotted against object distance. The two curves are extrapolated from data-fit functions⁵; their parameters are $y = 0.7658e^{-0.9249x}$, $r = -0.888$, $P = 0.0002$ (red curve) and $y = 0.9445e^{-0.6343x}$, $r = -0.928$, $P < 0.0001$. Values of r and P were obtained by using a linear transfer function method³⁰. The double arrow indicates the distance between curves when the sphere was 3 cm from the fish.

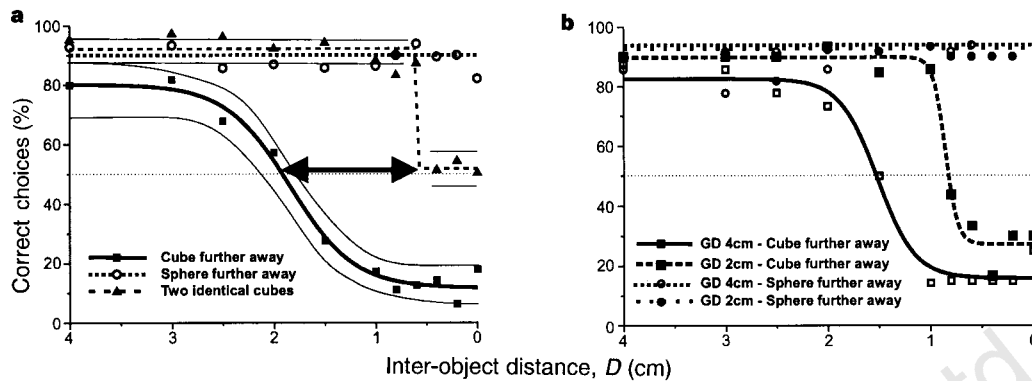


Figure 4 Psychometric functions for distance discrimination of a metal sphere and a metal cube, and for two identical metal cubes. **a**, Gate distance was fixed at 3 cm. Two curves are shown for the sphere/cube discrimination task: results obtained when either the cube (squares) or the sphere (circles) were further away. The double arrow marks the shift to the left of the curve fitted to the squares

bears some resemblance to the mechanism that flies use to land on a target: the onset of deceleration is triggered when the ratio of the image expansion of the target on the retina to the image size reaches a critical value²⁰. To be independent of object parameters, the image expansion parameter is standardized by image size. Similarly, in our hypothesis about active electric imaging, the slope parameter is standardized by maximal amplitude, making the slope-to-amplitude ratio independent of object parameters such as size, impedance or form.

In conclusion, we found that electric fish can measure the distance of objects in complete darkness by means of active electrolocation. Distance discrimination is very accurate, as long as the geometrical forms of the edges of the objects being compared do not differ too strongly. Fish can perceive three-dimensional depth without time or spectral measurements even when the receptive surface is stationary. □

Methods

Discrimination training. Three individuals of the weakly electric African fish *Gnathonemus petersii* (standard lengths between 11 and 14 cm) were trained in a rectangular tank that was divided into two compartments (60 cm × 35 cm and 20 cm × 35 cm) by a coarse plastic mesh containing two gates 15 cm apart (Fig. 1c). For the electrolocation discrimination task, two test objects were mounted in the smaller compartment behind the left and right gates at a distance chosen in a pseudo-random manner. The test objects were solid aluminium cubes (volumes 125, 64, 27 or 8 cm³), plastic cubes (125, 64, 27 or 8 cm³), aluminium plates (5 cm × 5 cm × 2 cm), aluminium pyramids (64 cm³) and aluminium spheres (diameter 4 cm or 2 cm). In a food-rewarded two-alternative forced-choice procedure, animals were trained to swim through that gate behind which the object was located farther away (S+). Distance-discrimination thresholds (defined as 70% correct choices) were measured by keeping the gate distance of the closer object constant, and by reducing the distance between the second object and its corresponding gate in progressive steps, thus reducing the distance between the two objects (D in Fig. 1c). Each point of the psychometric functions determined in this manner was based on at least 150 decisions. Except for controls, experiments were conducted at low light intensities (0.75 lx).

Control experiments. Three types of control experiment were conducted. (1) To measure the effect of the dim illumination, five experiments were conducted in infrared light, which is invisible to mormyrids²⁸. Threshold curves thus obtained in the absence of vision, and with light present, were compared. There was no statistical difference between thresholds obtained with or without visible light ($P < 0.05$). Thus, even when vision was possible, the fish relied on active electrolocation to determine object distance. (2) To prove that fish were using active electrolocation for the required task, electrical noise (bandwidth 2–5 kHz, minimum noise amplitude = 70 mV cm⁻¹ (r.m.s.)) was played to the fish during the discrimination task. The two fish tested were unable to

relative to that obtained with two identical cubes (triangles). Thin lines indicate 95% confidence limits for the fitted functions. All curves are significantly different ($P < 0.05$). **b**, Psychometric functions obtained when a fish discriminated between a sphere and a cube at a gate distance of 2 cm (squares) or 4 cm (circles).

discriminate object distances under these conditions and reached success rates of $50 \pm 10\%$. (3) To prove that fish were not making a comparison by electrolocating the two objects simultaneously, a thin glass plate that was impenetrable to electrolocation signals was mounted between the two objects, making it impossible for the fish to image both at the same time. The performances of the fish were statistically identical ($P < 0.05$) with and without the glass divider present. Psychometric functions were compared by transforming the sigmoid curves into linear functions, recalculating data points according to this transfer function and comparing slopes with a modified Student's *t*-test as described by Zar²⁹.

Measurement of electric images. The electric images of the same objects that were used in the behavioural experiments were measured experimentally along the lateral electrosensitive surface of the fish (Fig. 3a). Three *G. petersii* were sedated with Hypnodil (metomidate R7315; Janssen Veterinary Division; 5 mg l⁻¹ aquarium water) and supported in the middle of a 40 cm × 20 cm recording tank. Under Hypnodil, a fish discharges its electric organ at a regular slow rhythm. Electric organ discharges (EODs) were recorded with a pair of copper wire electrodes insulated except at the tip (distance between electrodes, 1 mm). To avoid capacitive coupling, these electrodes were individually shielded and connected separately to the two inputs of an Axoclamp 2A amplifier. The two outputs of the amplifier were connected to a differential amplifier (Tektronix 5A22N) in a Tektronix 5223 digitizing oscilloscope. The recorded EOD signal was transferred to a computer hard disc by using ACQUIS1 software (G. Sadoc, C.N.R.S.). The electrode pair was moved in a straight line, in steps of 3 mm, along the side of the fish at the height of the lateral line, at a constant distance of 1 mm (5 mm in one fish) from the skin surface. At each position, ten records of the local EOD were averaged and the peak-to-peak amplitude of the EOD was determined. Control measurement series were made in the absence of any object between each trial series. Images of several types of objects located at seven different distances from the fish's surface were measured. The relative modulation of the EOD peak-to-peak amplitude was defined as the ratio between the change produced by the object and the amplitude in the absence of an object. The electric image of an object was defined as the profile of the relative modulation along the fish's body (Fig. 3b)⁷. For each electrical image, we determined the maximal image amplitude and maximal image slope values. The maximal value of the rostral slope of the image was divided by the maximal image amplitude to obtain the slope-to-amplitude ratio for each image.

Received 16 June; accepted 20 August 1998.

- Collett, T. S. & Harkness, L. I. K. in *Analysis of Visual Behaviour* (eds Ingle, D. J., Goodale, M. A. & Mansfield, J. W.) 111–175 (MIT Press, 1982).
- Lissmann, H. W. & Machin, K. E. The mechanism of object location in *Gymnarchus niloticus* and similar fish. *J. Exp. Biol.* **35**, 451–486 (1958).
- von der Emde, G. in *The Physiology of Fishes* (ed. Evans, D. H.) 313–343 (CRC, Boca Raton, 1998).
- Bastian, J. in *Comparative Perception* (eds Stebbins, W. C. & Berkeley, M. H.) 35–89 (Wiley, New York, 1989).
- Schwarz, S. Diplomarbeit, Institut für Zoologie, Universität Bonn (1997).
- Heiligenberg, W. Electrolocation of objects in the electric fish *Eigenmannia* (Rhamphichthyidae, Gymnotoidei). *J. Comp. Physiol.* **87**, 137–164 (1973).
- Caputi, A., Budelli, R., Grant, K. & Bell, C. C. The electric image in weakly electric fish. II Physical images of resistive objects in *Gnathonemus petersii*. *J. Exp. Biol.* **201**, 2115–2128 (1998).

8. Rasnow, B. The effects of simple objects on the electric field of *Apteronotus*. *J. Comp. Physiol. A* **178**, 397–411 (1996).
9. Howard, I. P. & Rogers, B. J. *Binocular Vision and Stereopsis* (Oxford Univ. Press, New York, 1995).
10. Collett, T. Stereopsis in toads. *Nature* **267**, 349–351 (1977).
11. Kral, K. & Poteser, M. Motion parallax as a source of distance information in locusts and mantids. *J. Insect Behav.* **10**, 145–163 (1997).
12. Lehrer, M., Wehner, R. & Srinivasan, M. V. Motion cues provide the bee's visual world with a third dimension. *Nature* **332**, 356–357 (1988).
13. Schuij, A. & Hawkins, A. D. Acoustic distance discrimination by the cod. *Nature* **302**, 143–144 (1983).
14. Bleckmann, H. *Reception of Hydrodynamic Stimuli in Aquatic and Semiaquatic Animals* (Gustav Fischer, Stuttgart, 1994).
15. Atema, J. Eddy chemotaxis and odor landscapes: exploration of nature with animal sensors. *Biol. Bull.* **191**, 129–138 (1996).
16. Moller, P., Serrier, J., Belbenoit, P. & Push, S. Notes on the ethology and ecology of the Swashi river mormyrids (Lake Kainji, Nigeria). *Behav. Ecol. Sociobiol.* **4**, 357–368 (1979).
17. von der Emde, G. The sensing of electric capacitances by weakly electric mormyrid fish: effects of water conductivity. *J. Exp. Biol.* **181**, 157–173 (1993).
18. Tippler, P. A. *Physik* (Spektrum Akademischer, Heidelberg, 1994).
19. von der Emde, G. & Bleckmann, H. Finding food: senses involved in foraging for insect larvae in the electric fish, *Gnathonemus petersii*. *J. Exp. Biol.* **201**, 969–980 (1998).
20. Wagner, H. Flow-field variables trigger landing in flies. *Nature* **297**, 147–148 (1982).
21. Hassan, E. S. in *The Mechanosensory Lateral Line. Neurobiology and Evolution* (eds Coombs, S., Gerner, P. & Münz, H.) 217–228 (Springer, Berlin, 1989).
22. Bleckmann, H., Tittel, G. & Blübaum-Gronau, E. in *The Mechanosensory Lateral Line. Neurobiology and Evolution* (eds Coombs, S., Gerner, P. & Münz, H.) 501–526 (Springer, Berlin, 1989).
23. Harkness, L. Chameleons use accommodation cues to judge distance. *Nature* **267**, 346–349 (1977).
24. Hopkins, C. D. Temporal structures of non-propagated electric communication signals. *Brain Behav. Evol.* **28**, 43–59 (1986).
25. Suga, N., Butman, J. A., Teng, H., Yan, J. & Olsen, J. F. in *Active Hearing* (eds Flock, A., Ottoson, D. & Ulfendahl, M.) 13–30 (Pergamon, New York, 1995).
26. Schnitzler, H.-U., Menne, D. & Hackbarth, H. in *Time Resolution in Auditory Systems* (ed. Michelsen, A.) 180–204 (Springer, Berlin, 1985).
27. Dear, S. P., Simmons, J. A. & Fritz, J. A possible neuronal basis for representation of acoustic scenes in auditory cortex of the big brown bat. *Nature* **364**, 620–623 (1993).
28. Cialo, S., Gordon, J. & Moller, P. Spectral sensitivity of the weakly discharging electric fish *Gnathonemus petersii* using its electric organ discharges as the response measure. *J. Fish Biol.* **50**, 1074–1087 (1997).
29. Zor, J. H. *Biostatistical Analysis* (Prentice Hall, Englewood Cliffs, 1984).
30. Hoerl, A. E. Jr in *Chemical Business Handbook* (ed. Perry, J. H.) 20–50 (McGraw-Hill, London, 1954).

Acknowledgements. The behavioural experiments and the analysis of the electric images were performed by S.S. during work for his diploma thesis. We thank H. Bleckmann for providing laboratory space and for his continuous support throughout this study; C. C. Bell, H. Bleckmann, J. Mogdans, S. F. Perry, F. Schaeffel and H. Wagner for critically reading the manuscript and helpful discussions; C. Gutzen for help with the figures; and W. Alt for statistical advice. G.v.d.E. is a recipient of a Heisenberg stipend of the DFG. This work was financed partly by a research grant from the EC to K.G. by the Franco-German international exchange programme PROCOPE, by the Franco-Uruguayan exchange programme ECOS and by a doctoral fellowship to L.G. from the French Ministry of Foreign Affairs.

Correspondence and requests for materials should be addressed to G.v.d.E. (e-mail: unb308@uni-bonn.de).

Seeing biological motion

Peter Neri^{*}, M. Concetta Morrone[†] & David C. Burr[‡]

^{*} University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK

[†] Istituto di Neurofisiologia del CNR, Via S. Zeno 51, Pisa 56127, Italy

[‡] Department of Psychology, University of Florence, Florence 50125, Italy

One of the more stunning examples of the resourcefulness of human vision is the ability to see 'biological motion', which was first shown¹ with an adaptation of earlier cinematic work²: illumination of only the joints of a walking person is enough to convey a vivid, compelling impression of human animation, although the percept collapses to a jumble of meaningless lights when the walker stands still. The information is sufficient to discriminate the sex and other details of the walker^{3,4}, and can be interpreted by young infants⁵. Here we measure the ability of the visual system to integrate this type of motion information over space and time, and compare this capacity with that for viewing simple translational motion. Sensitivity to biological motion increases rapidly with the number of illuminated joints, far more rapidly than for simple motion. Furthermore, this information is summed over extended temporal intervals of up to 3 seconds (eight times longer than for simple motion). The steepness of the summation curves indicates that the mechanisms that analyse biological motion do not integrate linearly over space and time with constant efficiency, as may occur for other forms of complex motion⁶, but instead adapt to the nature of the stimulus.

Biological motion was produced on a video display by an adaptation of a standard cyclic algorithm⁷ that emulates the

motion of the 11 main joints of a person walking on a treadmill. Subjects were required either to detect the presence of the walker or to discriminate the direction of walking, in the presence of variable amounts of dynamic random noise. Similar tasks were performed for simple translational motion, for which a random pattern of dots was continuously displaced horizontally within a window of dimensions similar to that of the walker. We adapted the standard biological-motion technique so that each point had a 'limited lifetime' of two frames, after which it disappeared and was redrawn in another randomly chosen position (see Fig. 1 and Methods). This allowed for dynamic undersampling of the entire dot pattern, ensuring that, even at the lowest sampling rates, all 11 joints were likely to be represented during the 1,200-ms exposure time.

Figure 2 shows how sensitivity (expressed as maximum tolerable noise) for detection and direction discrimination varied with the number of displayed points for the two types of motion. For translational motion, sensitivity for both detection and discrimination increased with dot number at a very similar rate, almost linearly for both subjects tested (log-log slope is roughly at unity). For biological motion, the summation curves for detection were very similar to those for detection of translation, in both slope and absolute sensitivity. But the direction-discrimination curves were far steeper, with log-log slopes of about 3 or more (indicating a cubic or higher relationship).

Figure 3 shows how sensitivity to the walker varied with exposure time (temporal summation). As before, we used a limited-lifetime paradigm, with six dots displayed. Sensitivity to both simple and biological motion increased with time, first rapidly then more gradually. In comparable studies, it was thought that the initial rapid increase reflects physiological summation, whereas the gradual phase reflects 'probability summation' (a statistical improvement based on recruitment of independent detectors rather than summation within a single unit⁸). The intersection of these two curves gives an estimate of the time constant of the

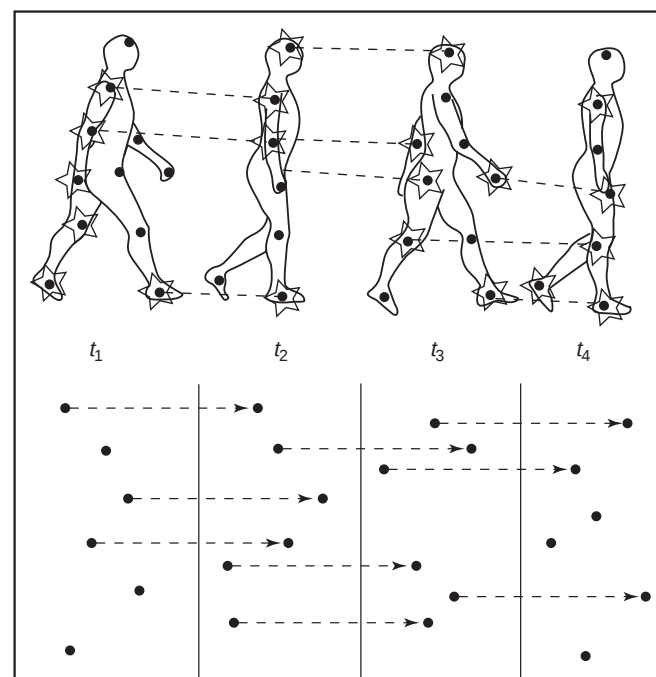


Figure 1 The limited-lifetime technique for studying the ability to see biological motion (upper figures) and translational motion (lower figures), for six-dot sampling (over a two-frame running average). The starred points of the walker indicate those that are actually in motion; the others, the possible positions to be sampled. Half of the dots move from frames 1 to 2, and the other half from frames 2 to 3. The starred points sometimes undergo occlusion (when required by the algorithm⁷), and are not visible at these times.

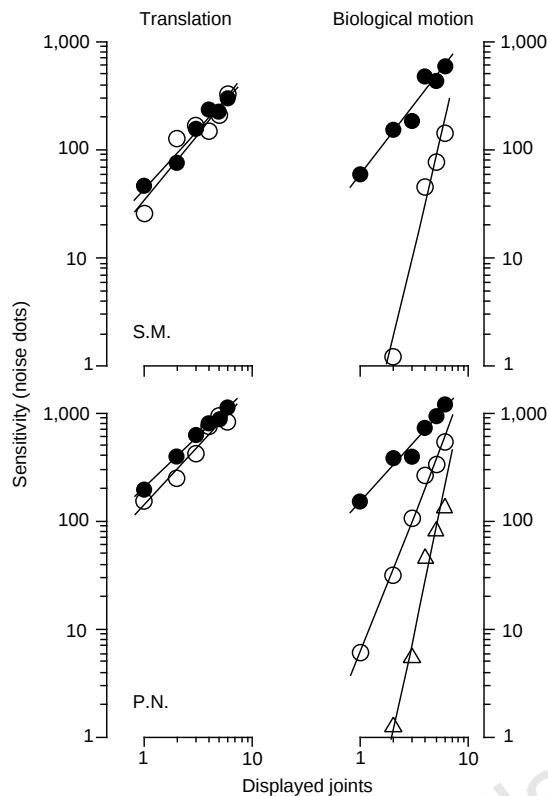


Figure 2 Sensitivity, expressed as the number of noise dots at threshold, for detection (filled circles) and direction discrimination (open circles), for both simple translation and biological motion. The open triangles indicate results of a study in which the observer was required to discriminate 'coherent' from 'incoherent' motion; in this case, the upper and lower body moved either in the same or in opposite directions²⁰. The steepness of the log-log curves (calculated by linear regression) for subjects S.M. and P.N. (respectively) were: 1.10 and 0.96 for translation detection; 1.24 and 1.08 for translation discrimination; 1.29 and 1.12 for detection of biological motion; 4.23 and 2.55 for discrimination of biological motion; and 4.48 for discriminating coherent from incoherent motion (P.N.). S.M. was naïve of the goals of the study.

physiological summation along local trajectories⁹. For simple translation, the time constant of summation was ~ 600 ms, and was nearly invariant with velocity. For biological motion, however, the time constants were much larger, up to 2,800 ms. Furthermore, the estimate of summation depended on image velocity, in roughly inverse proportion, indicating that the limiting factor was the number of cycles presented rather than time passed *per se*.

The results indicate several new facts about biological motion. They demonstrate objectively the robustness of this curious stimulus; the visual system is very sensitive to fragments of natural motion scenes: observers reliably detected the direction of motion of point-lit figures in greatly degraded conditions, when half of the points were of opposite polarity, local-motion trajectories were given by two-frame sequences, and up to 1,000 similar dots were sprinkled over the limited area. Indeed, with adequate spatial sampling, sensitivity for discriminating the direction of ambulation was as high as that for its detection, or for detection of discrimination of direction of simple translation. Temporal summation for biological motion occurs over much longer periods than for simple translation, and seems to be limited by the number of cycles of ambulation, rather than by total duration of the stimulus.

Theoretically, sensitivity might be expected to increase linearly with the spatial sampling density¹⁰, as observed in discrimination of Glass patterns^{11,12}, simple translational motion^{13,14} and complex optic-flow patterns^{6,15}. The same relationship occurred here for detection and discrimination of translation, and for detection of

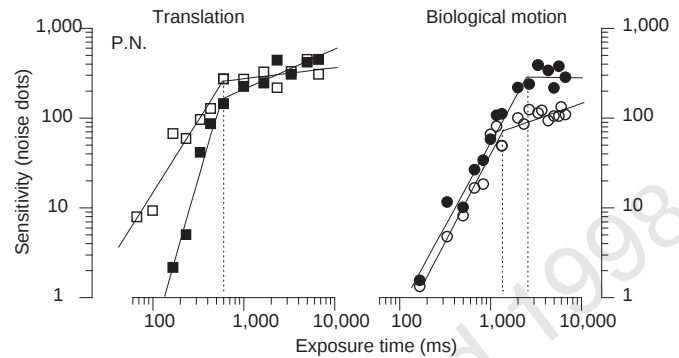


Figure 3 Sensitivity for direction discrimination, for both translation and biological motion. The following range of image speeds was used: 1.3 deg s⁻¹ (filled squares) and 7.3 deg s⁻¹ (open squares) for translation; and 1.3 deg s⁻¹ (filled circles) and 3.8 deg s⁻¹ (open circles) for biological motion, using the limited-lifetime procedure (with six joints) (see Methods). The dashed vertical lines indicate the transition from physiological to probability summation, calculated by best-fit of a two-segment template of variable slope. The summation constant for translation was relatively short (600 ms) and invariant with speed, whereas summation for biological motion was longer and varied inversely with image speed. The initial steep slope of ~ 2 for all curves is predicted by a linear integrator given that the stimulus was embedded in a 7-s noise sequence. We obtained similar results with another subject (S.M.).

biological motion. This latter result may reflect independent integration of signals along the trajectory of each joint by local-motion mechanisms, given that the sampling rate of each trajectory will increase directly with the sampling rate of the joints, and that detection of a single joint is sufficient to reveal the location of the walker.

Direction and coherence discrimination of biological motion varied more with undersampling than did detection, so, at low sampling rates, direction could not be discerned at noise levels that were nearly two logarithmic units below the detection threshold. If there were a linear integrator (or matched filter) that summates the motion signal along all joint trajectories, then sensitivity for direction and coherence discrimination should increase linearly with sampling rate, as it does for the examples mentioned above. The far steeper slope observed for biological motion may indicate that these types of task are mediated by detectors with efficiency that varies dynamically with the signal strength. It has been suggested that biological motion depends on virtual links between appropriate joints, requiring the simultaneous analysis of both joints^{1,16-18}. This idea is intuitively appealing, but the requirement of coincidence does not in itself predict the steep summation, because it affects both signal and noise equally, so the theoretical prediction is still a curve of unit slope (or an even shallower slope, depending on the assumptions about the coincidences produced by random noise).

Our results indicate that there is probably no specialized hardware for biological motion that acts like an ideal detector, as seems to exist for other forms of simple and complex motion or spatial vision^{6,11-15}. Instead, the results show that biological motion may be analysed by very sensitive, but flexible, mechanisms with variable efficiency. One method of achieving variable efficiency would be by a system that adjusts its internal noise (or threshold) to respond optimally to the stimulus under all conditions. The advantage of such a system over the seemingly simpler matched-filter systems may be that it can optimize a limited neural resource for the analysis of the much wider range of stimuli that can yield information about spatial structure from motion. □

Methods

Both biological and translational motion were created with a 'limited-lifetime' technique, at a sampling rate of 30 Hz (Fig. 1). Each signal dot moved to the next position in the motion sequence, then disappeared; it was 'reborn' at a

joint chosen randomly from those not sampled in the previous sequence. All signal and noise dots were 7 arcmin in diameter, and were half-white and half-black against a grey background of 20 cd m⁻² (thus causing no change in mean luminance). The walker was generated by Cutting's algorithm⁷ (with no net translation) from a randomly chosen starting position, usually at 0.75 gait-cycles per second, with the individual dots moving at an average speed of 1.8 deg s⁻¹. For translation, dots were placed in random positions over the area of the walker, and all moved at 1.8 deg s⁻¹ (to obtain results shown in Fig. 2). The walker was symmetrical about the vertical midline (particularly the upper body), so spatial cues alone could not aid discrimination.

Observers fixated the centre of a Barco Calibrator monitor (frame rate 180 Hz) from a distance of 80 cm. After a warning signal, the target was presented to either the left or the right of fixation, with an appropriate density control on the other side (both stimuli subtended 3.4 × 5.7°, centred 2.4° from fixation). For the walker, the density control was derived from the walking algorithm by randomizing the order of the frames presented. For translation, the control dots were displayed in new random positions within the region on each frame. Thus both target and control were dynamic, but only in the target was the motion coherent and smooth. Detection of either class of stimuli could be based on a judgement of smoothness of motion of individual dots. Discrimination was a two-stage process, in which observers first selected which side contained the target, and then identified the direction of the moving dots (for translation), the direction of ambulation (for discrimination of biological motion), or whether the upper and lower body of the walker moved coherently. As the discrimination thresholds for translation were similar to those for detection, it is unlikely that the two-stage task was an impediment to performance.

Dynamic random noise, comprising dots of similar size and colour, was scattered over the entire screen. The density of the noise increased or decreased in each trial, depending on the correctness of the observer's response (following the adaptive procedure QUEST¹⁹, without feedback). There were 200–400 trials for each condition, with sensitivity defined as the noise level at which 75% correct responses are made; sensitivity was calculated by fitting a raised cumulative gaussian curve (with asymptotes at 0.5 and 1) to the psychometric functions. For spatial summation (Fig. 2), stimuli were presented for 1,200 ms (almost one complete gait-cycle, comprising 40 frames) centred within a noise window of 1,400 ms. For temporal summation (Fig. 3), the stimulus interval varied within a noise window of 7 s.

Received 12 June; accepted 21 September 1998.

- Johansson, G. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* **14**, 201–211 (1973).
- Marey, E.-J. *Le Mouvement* (Masson, Paris, 1894).
- Mather, G. & Murdoch, L. Gender discrimination in biological motion displays based on dynamic cues. *Proc. R. Soc. Lond. B* **259**, 273–279 (1994).
- Dittrich, W. H., Troscianko, T., Lea, S. & Morgan, D. Perception of emotion from dynamic point-light displays represented in dance. *Perception* **25**, 727–738 (1996).
- Fox, R. & McDaniel, C. The perception of biological motion by human infants. *Science* **218**, 486–487 (1982).
- Morrone, M. C., Burr, D. C. & Vaina, L. Two stages of visual processing for radial and circular motion. *Nature* **376**, 507–509 (1995).
- Cutting, J. A program to generate synthetic walkers as dynamic point-light displays. *Behav. Res. Methods Instrument* **10**, 91–94 (1978).
- Watson, A. B. Probability summation over time. *Vision Res.* **19**, 515–522 (1979).
- Burr, D. C. Temporal summation over moving images by the human visual system. *Proc. R. Soc. Lond. B* **211**, 321–339 (1981).
- Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
- Maloney, R. K., Mitchison, G. J. & Barlow, H. B. Limit to the detection of Glass patterns in the presence of noise. *J. Opt. Soc. Am. A* **4**, 2336–2341 (1987).
- Wilson, H. R., Wilkinson, F. & Asaad, W. Concentric orientation summation in human form vision. *Vision Res.* **37**, 2325–2330 (1997).
- Lappin, J. S. & Bell, H. H. The detection of coherence in moving random-dot patterns. *Vision Res.* **16**, 161–168 (1976).
- Barlow, H. B. & Tripathy, S. P. Correspondence noise and signal pooling in the detection of coherent visual motion. *J. Neurosci.* **17**, 7954–7966 (1997).
- Burr, D. C., Morrone, M. C. & Vaina, L. Large receptive fields for optic flow direction in humans. *Vision Res.* **38**, 1731–1743 (1998).
- Cutting, J. E. Coding theory adapted to gait perception. *J. Exp. Psychol.* **7**, 71–87 (1981).
- Marr, D. & Vaina, L. Representation and recognition of the movements of shapes. *Proc. R. Soc. Lond. B* **214**, 501–524 (1982).
- Ullman, S. in *Human and Machine Vision* (eds Beck, J., Hope, B. & Rosenfeld, A.) 459–480 (Academic, New York, 1983).
- Watson, A. B. & Pelli, D. G. QUEST: A Bayesian adaptive psychometric method. *Percept. Psychophys.* **33**, 113–120 (1983).
- Mather, G., Radford, K. & West, S. Low-level visual processing of biological motion. *Proc. R. Soc. Lond. B* **249**, 149–155 (1992).

Acknowledgements. We thank H. Barlow and J. Ross for useful discussions. P.N. was supported by a scholarship from the Scuola Normale Superiore, Pisa. Supported by MURST and EC BIOMED (VIPROM).

Correspondence and request for materials should be addressed to D.C.B. (e-mail: daye@in.pi.cnr.it).

Temporal dynamics of chromatic tuning in macaque primary visual cortex

Nicolas P. Cottaris* & Russell L. De Valois*†

* Program in Vision Science and † Department of Psychology, University of California Berkeley, 3210 Tolman Hall, California 94720, USA

The ability to distinguish colour from intensity variations is a difficult computational problem for the visual system because each of the three cone photoreceptor types absorb all wavelengths of light, although their peak sensitivities are at relatively short (S cones), medium (M cones), or long (L cones) wavelengths. The first stage in colour processing is the comparison of the outputs of different cone types by spectrally opponent neurons in the retina and upstream in the lateral geniculate nucleus^{1–3}. Some neurons receive opponent inputs from L and M cones, whereas others receive input from S cones opposed by combined signals from L and M cones. Here we report how the outputs of the L/M- and S-opponent geniculate cell types are combined in time at the next stage of colour processing, in the macaque primary visual cortex (V1). Some V1 neurons respond to a single chromatic region, with either a short (68–95 ms) or a longer (96–135 ms) latency, whereas others respond to two chromatic regions with a difference in latency of 20–30 ms. Across all types, short latency responses are mostly evoked by L/M-opponent inputs whereas longer latency responses are evoked mostly by S-opponent inputs. Furthermore, neurons with late S-cone inputs exhibit dynamic changes in the sharpness of their chromatic tuning over time. We propose that the sparse, S-opponent signal in the lateral geniculate nucleus is amplified in area V1, possibly through recurrent excitatory networks. This results in a delayed, sluggish cortical S-cone signal which is then integrated with L/M-opponent signals to rotate the lateral geniculate nucleus chromatic axes^{4–5}.

The term 'receptive field' is used traditionally to characterize how a neuron responds to stimuli in different spatial locations, thus referring to a spatial receptive field⁶. Here we are concerned with determining how a cortical neuron responds to stimuli consisting of chromatic shifts from a white point to different locations in colour space, and how this responsivity develops over time. We are therefore studying a chromatic–temporal receptive field. To examine the cortical transformation of lateral geniculate nucleus (LGN) chromatic signals directly, we specified stimulus chromaticity in the MacLeod–Boynton–Derrington–Krauskopf–Lennie (MBDKL) isoluminant plane^{7,8}. This plane is defined by two axes whose chromaticities isolate responses from the two types of LGN opponent neuron⁸: the 0° to 180° axis isolates responses from L/M-opponent LGN neurons (0°: L – M, 'pinkish-red'; 180°: M – L, 'cyan'); and the 90° to 270° axis isolates responses from S-opponent LGN neurons (90°: S – (L + M), 'violet'; 270°: – S + (L + M), 'greenish-yellow'). A cortical cell that integrates signals from both types of colour-coding geniculate cell, as most cortical cells do, would have a preferred chromaticity at an intermediate angle, depending on the relative weights and timings of the cell's inputs.

To study the structure of the chromatic–temporal receptive field, we probed neurons with spatially uniform chromatic stimuli presented in a fast (30-ms flashes), pseudorandom sequence and analysed the neuronal response using the reverse-correlation procedure^{9,10}. Stimulus dimensions were one to two times the classical receptive-field dimensions, as measured with *m*-sequence receptive-field mapping¹¹. For each action potential (spike) fired, we determined which chromatic stimulus had been presented at various preceding times. Spikes were accumulated in a two-dimensional

Table 1 Laminar distribution of cells

Chromatic-temporal receptive field	2 + 3	4A/4B	4C _{α+β}	5	6	Unknown	Total
Early, single-peaked	5	5	12	1	11	6	40
Late, single-peaked	11	3	4	4	3	0	25
Double-peaked	6	6	6	4	3	0	25
Continuously shifting	0	1	0	0	4	0	5
Cells per layer	22	15	22	9	21	6	95

Number of cells of various chromatic-temporal dynamics found in each cortical layer. More than half of the cells with early-peaking receptive fields were found in layers 4C and 6 and more than half of the cells with late-peaking receptive fields were found in the supragranular layers (2 + 3 and 4A/4B). Cells with double-peaked receptive fields were found in all layers. Almost all cells with continuously shifting receptive fields were found in layer 6.

histogram corresponding to stimulus chromaticity and time delay before spiking. After some thousands of spikes were processed this way, we found that, immediately before spiking, all stimuli were equally likely to have been presented as it takes around 30–50 ms for visual information to reach cortical area V1. However, for longer time delays, chromatic stimuli that excite the neuron would be more likely to have been presented, and those that inhibit the neuron would be less likely to have been presented.

The data set consists of 95 cells from 12 monkeys, all of which responded reliably and consistently to our rapidly changing stimulus (see Methods). About two-thirds of the cells showed a single chromatic preference over time, indicating that their different chromatic inputs all had similar temporal dynamics, whereas the remaining one-third of the cells exhibited major shifts in their colour tuning over time, reflecting multiple chromatic inputs with different latencies. Cells with a single chromatic peak appeared to fall into two classes on the basis of their latencies between stimulus presentation and spiking, namely those with early latencies (<96 ms) and those with longer latencies (≥96 ms). The validity of this division was reinforced by finding that these two classes differed also in their chromatic preferences and their chromatic-integration dynamics.

Neurons with early, single-peaked chromatic-temporal receptive fields (40 of 95) were found mainly in cortical layers 4C and 6 (Table 1). Their responses peaked between 68 and 96 ms after stimulus presentation (Fig. 1A), and they maintained a nearly constant sharpness of chromatic tuning during their response course (see below and Fig. 2), indicating static and instantaneous integration of their chromatic inputs. A smaller number of these neurons (3 of 40) had very short times-to-peak (55–70 ms), biphasic temporal impulse responses, strong inverted S-cone inputs (Fig. 1A, right-most receptive field) and were found in layer 6.

Neurons with late, single-peaked chromatic-temporal receptive fields (25 of 95) were found mainly in the supragranular layers (14 of 25). Their responses peaked between 96 and 135 ms after stimulus presentation and they had slower dynamics than the early-peaking neurons (Fig. 1B). In addition, the sharpness of chromatic tuning in most of these cells changed over time (Fig. 2c, inset), suggesting dynamic integration of their chromatic inputs. Almost all of these neurons had strong S-cone inputs with peak chromatic axes within 30° of the ±S geniculate axis.

Neurons with major shifts in their chromatic tuning over time were of two types, namely those with double-peaked receptive fields, exhibiting shifts of around 90° (25 of 95), and those exhibiting continuous 360° chromatic shifts (5 of 95). Most double-peaked neurons (22 of 25) responded first to stimuli around the L–M chromatic axis, followed 10–30 ms later by responses to stimuli around the S axis (Fig. 1Ca). The reverse configuration, early responses to S-stimuli and late responses to L/M-stimuli, was observed in 3 of 25 neurons (Fig. 1Cb). Double-peaked neurons were found in all layers. Neurons with continuously shifting receptive fields (Fig. 1D) started responding very early (40–50 ms after stimulus presentation), and had biphasic temporal impulse responses and significant S-cone inputs. Their chromatic-tuning sharpness at any point in time was nearly constant. Like the very

fast, biphasic neurons with strong inverted S-cone inputs (Fig. 1A, right-most receptive field; see above), these neurons were found mainly in layer 6 and may correspond to a postulated fast S-cone input to luminance¹² and motion-sensitive¹³ mechanisms.

To quantify our results we extracted the time-to-peak of a cell's different chromatic responses by curve-fitting separate receptive-field time slices. Figure 3a shows the distribution of the time-to-peak parameter compiled across all chromaticities (grey histogram) and across chromaticities within ±30° from the L/M- and the S-opponent axes (red and purple histograms, respectively). The trough around 96 ms in the aggregate distribution indicates that there might be two major cortical chromatic-response types, those peaking before and those peaking after 96 ms (early and late responses, respectively). Furthermore, striate responses evoked by L/M-opponent inputs are fast (median 81 ms, $\sigma = 7.1$ ms), comprising most of the early peak of the aggregate latency distribution, whereas responses evoked by signals from S-opponent inputs are slower (median 100 ms, $\sigma = 16.9$ ms), contributing about one-third of the early portion and most of the late portion of the aggregate time-to-peak distribution. These data concur with visual-evoked-potential (VEP) measurements that show longer latencies in response to S-cone-isolating stimuli¹⁴, although the neural origin of the VEP signal is unknown.

Figure 3b illustrates the relationship between receptive-field peak chromaticity and receptive-field time-to-peak for neurons with single-peaked receptive fields. These parameters were extracted from fits of the entire receptive field (Methods). The preferred chromaticities of early-peaking neurons are distributed throughout most of the 360° chromatic range, with subtle gaps between the geniculate axes, whereas late-peaking neurons are mostly tuned near the ±S-cone axis. On the other hand, nearly all double-peaked neurons (Fig. 3c) have their early chromatic peaks in response to L/M inputs, and their delayed peaks in response to S inputs. This is further evidence for longer processing delays of S- versus L/M-opponent signals, and indicates that the integrated output of double-peaked neurons would have preferred chromaticities distributed between the L/M and S axes; these chromaticities are under-represented by neurons with single-peaked receptive fields.

The difference in processing delays of L/M- and S-opponent signals may originate precortically, as signals from the two cone-opponent systems are processed independently in the retina¹⁵ and in the LGN^{8,16}. However, electrophysiological results show that L/M- and S-opponent LGN neurons have nearly identical temporal dynamics¹⁷. Examination of the shape of chromatic tuning as a function of time gives further support for a cortical origin of the delayed, sluggish S-cone signal we have identified. Our stimuli produce sinusoidal variation in L-, M-, and S-cone contrasts as a function of chromatic angle, and parvocellular LGN neurons, which sum cone signals linearly, have sinusoidal chromatic tuning in this colour space⁸. However, cortical static nonlinearities, such as response threshold¹⁸ and expansive relationships^{18,19} between membrane potential and spike generation²⁰, are expected to yield narrower chromatic tuning in cortical neurons. In addition, cortical recurrent, excitatory networks²¹ could dynamically change the sharpness of chromatic tuning.

We studied the sharpness and the dynamics of sharpening of

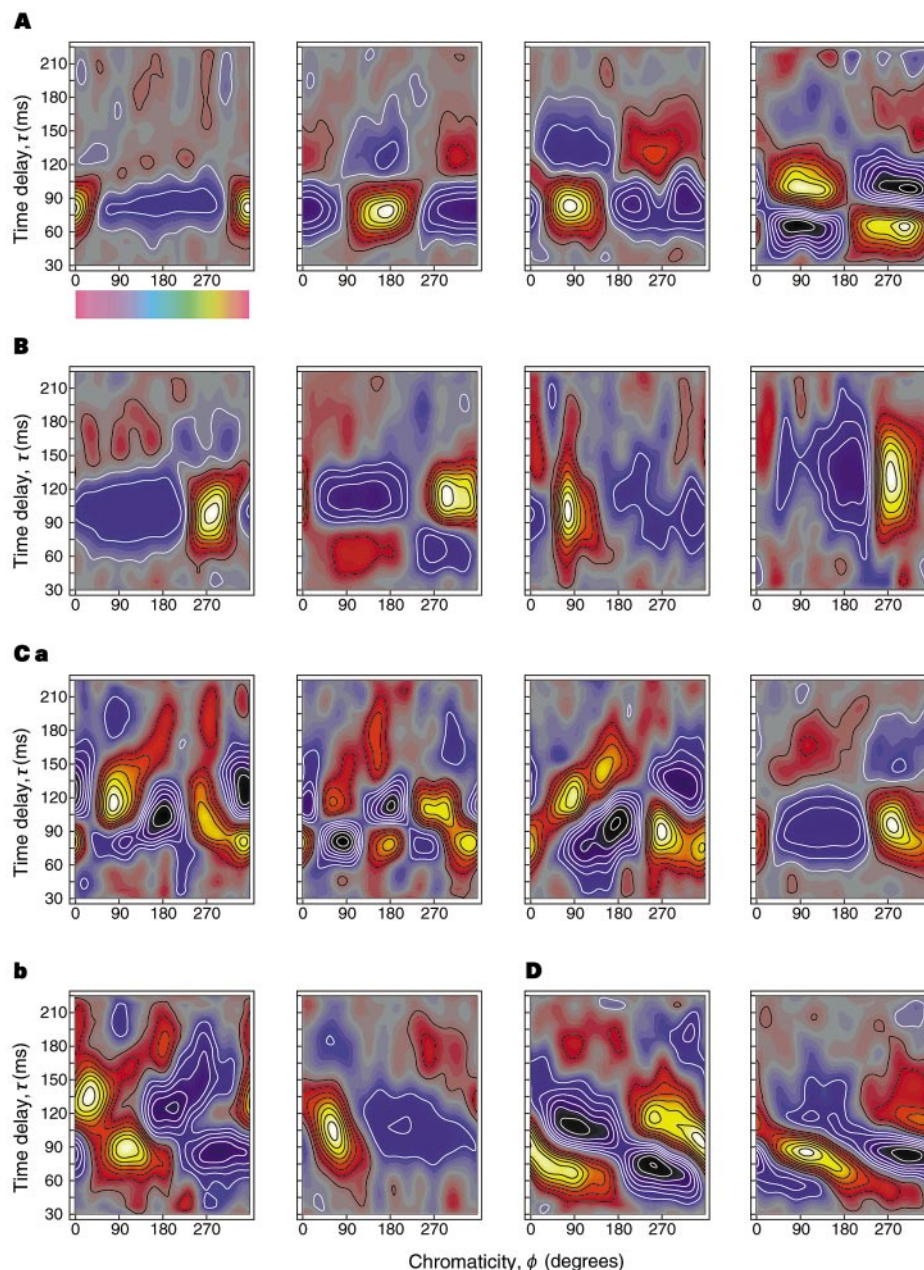


Figure 1 V1 chromatic-temporal receptive-field maps. X-axis: stimulus chromaticity (0° to 360°). Y-axis: response delay. Response magnitude is colour-coded (see Methods). Colour bar approximates the colour of stimuli. **A**, Neurons with early, single-peaked receptive field. Times-to-peak (left to right): 82 ms, 77 ms, 82 ms, and 62 ms. **B**, Neurons with late, single-peaked receptive field. Times-to-

peak (left to right): 97 ms, 111 ms, 98 ms, and 132 ms. **C**, Neurons with double-peaked receptive field. **a**, Early L/M-opponent input with late S-opponent input. **b**, Rare neurons with early S-opponent input and late L/M-opponent input, respectively. **D**, Neurons with continuously shifting chromatic tuning.

cortical chromatic tuning by fitting receptive-field slices parallel to the chromatic axis at successive time intervals with exponentiated sinusoidal functions with a threshold (Methods). Response exponents close to 1 indicate linear summation of cone signals (LGN-like), with progressively larger exponents indicating progressively sharper tuning. The chromatic temporal receptive fields of 74 out of 95 neurons could be fitted adequately with such functions. It was found that neurons with stable, early-peaking receptive fields had narrower colour tuning than LGN cells (mean exponent at peak, 1.57). Most of these neurons (30 of 39) had stable exponents over time (Fig. 2a), with the remainder exhibiting only small dynamic changes, indicating static nonlinearities as the mechanism of response sharpening. Neurons with continuously shifting chromatic tuning also showed stable exponents over time (Fig. 2b) with extremely linear summation properties (mean exponent at

peak, 0.97). In contrast, late-peaking neurons were very sharply tuned (mean exponent at peak, 4.67), and 15 out of 19 showed large variations in their response exponents over time (Fig. 2c), indicating dynamic nonlinearities as the mechanism of response sharpening. All but one of these 15 neurons had primarily $-S$ inputs. Changes in response exponent were also seen in the late, S -dominated peak in 9 out of 12 neurons with double-peaked receptive fields (Fig. 2d).

On the basis of these data, we propose that the relatively weak precortical S-cone signal (the signal is weak because of the sparse mosaic pattern of S cones²² and S-opponent ganglion²³ and LGN^{8,24} cells) is amplified in the cortex by a specialized mechanism, possibly a recurrent excitatory circuit in which the output of a striate neuron re-enters the cell to be summed synergistically with new signals. Lateral excitatory connections among proximal, S -dominated

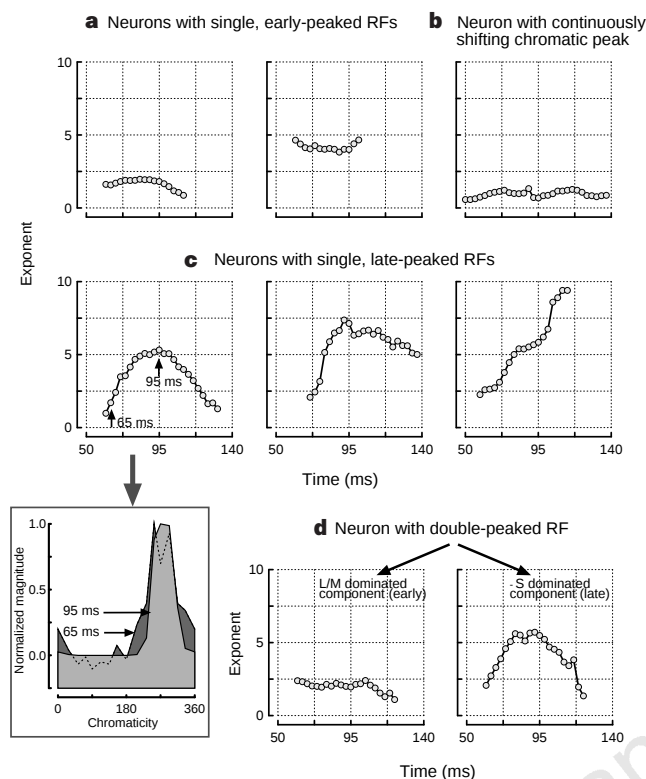


Figure 2 Dynamics of sharpening of V1 chromatic responses. Response exponent is plotted as a function of time. **a, b**, Neurons with non-changing exponents. **c**, Neurons with late-peaked receptive fields (RFs). Inset: chromatic tuning profile at 65 ms (dark grey) and at 95 ms (light grey). **d**, Neuron with double-peaked RF. This neuron's RF was fitted with the sum of two components whose chromaticities fluctuated around 330° (L/M-dominated) and 265° (S-dominated), which peaked at 80 ms and 98 ms, respectively. Of 33 neurons with dynamic changes in sharpness of chromatic tuning, the response exponent and response magnitude were correlated in 17. Response exponent increased or decreased continuously for 12 and 4 neurons respectively, of the remaining 16.

neurons may also contribute, as time-varying exponents, longer times-to-peak, and large S-cone responses are usually seen at very low spatial frequencies (N.P.C. and R.L.D.V., manuscript in preparation). The resulting S-cone signal is quite different from that seen in the LGN, having longer latency and slower dynamics, dynamically modified sharpness over time, and equal representation of both polarities (S cells are extremely rare in the retina and the LGN^{24,25}). This amplified, but sluggish, cortical S-cone signal becomes integrated with the abundant LGN L/M-opponent signal, rotating the LGN chromatic axes to form the perceptual 'red-green' and 'blue-yellow' colour axes⁵. This arrangement places a constraint in the minimal integration window needed by next-order neurons for the reliable sampling of signals from double-peaked neurons. Such integration windows would be of the order of 40–90 ms for signals from double-peaked neurons, contrasting with the 20–30-ms windows for signals from single-peaked neurons. This cortical amplification and nonlinear processing of the S-cone information concurs with recent psychophysical results²⁶. □

Methods

Rhesus monkeys (*Macaca mulatta*) were tranquilized with ketamine HCl (10–15 mg kg⁻¹, intramuscular). Anaesthesia was maintained with a continuous intravenous (i.v.) infusion of sufentanil citrate (5–12 µg kg⁻¹ h⁻¹, i.v.). Electrocardiogram, electroencephalogram, body temperature and expired CO₂ were continuously monitored. All the procedures were in accord with NIH guidelines and approved by the University of California Animal Care and

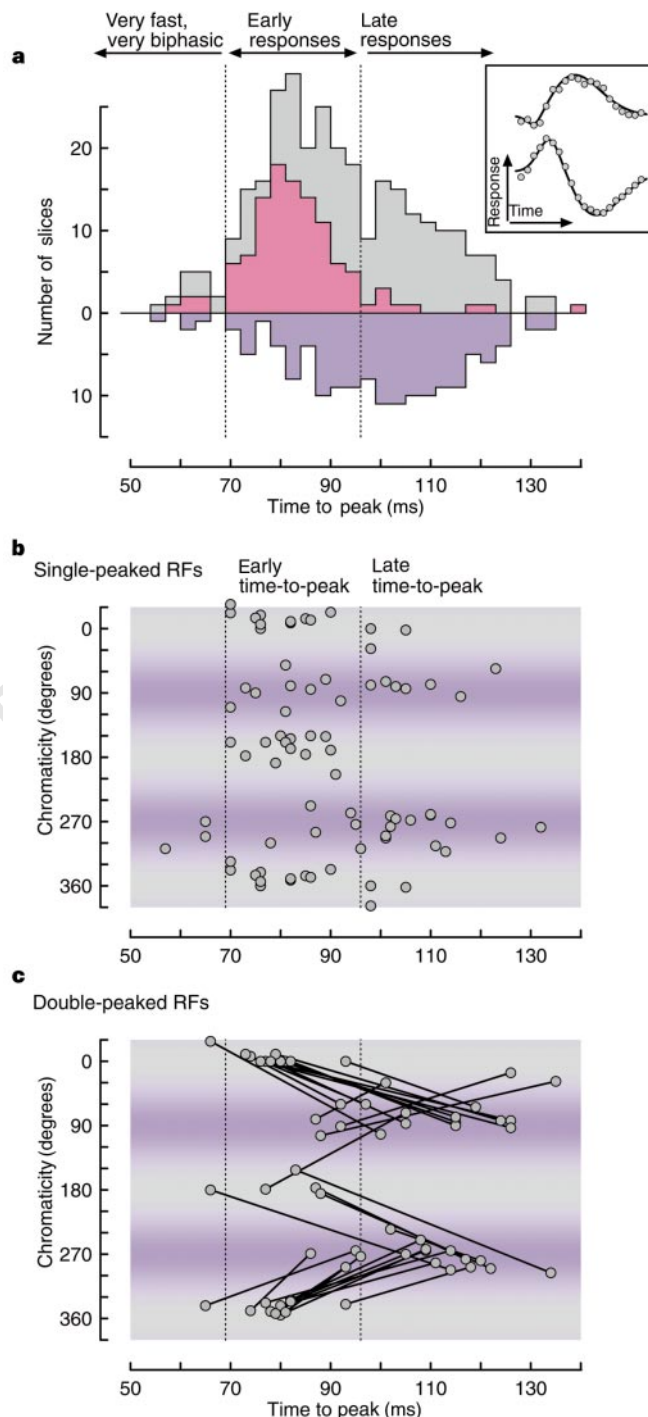


Figure 3 Times-to-peak of V1 chromatic responses. **a**, Distributions of times-to-peak extracted by curve fitting of separate receptive-field slices parallel to the time axis (see Methods). Grey, aggregate distribution compiled across all chromaticities (315 responses from 95 neurons); red and purple, distributions for slices taken at chromaticities within $\pm 30^\circ$ of the L-M and the S axes, respectively. Inset shows typical curve fits. **b**, Neurons with single-peaked receptive fields (RFs). RF peak chromaticity is plotted against its time-to-peak (see Methods). **c**, Neurons with double-peaked RFs. Each cell's characteristic 90° chromatic transition is sampled at two points (RF regions of peak response) and the points are connected to identify individual cells. Grey-purple grating in **b, c** depicts S-cone contrast variation.

Use Committee. Visual stimuli were generated by a SUN/TAAC-1 image processor, and displayed on an NEC monitor with a 66-Hz refresh rate and a mean luminance of 70 cd m⁻². Receptive fields were located within the central 5° . Action potentials were recorded with 1-ms resolution. Small lesions (5–10 µA × 5 s) were made every 500 µm along the length of each penetration.

Reconstruction of penetrations and assignment of cells to cortical layers was done using published criteria²⁷.

Colour space. Chromatic tuning was measured in the isoluminant plane of the three-dimensional MBDKL colour space^{7,8}, at chromaticity angles, ϕ , of 0°, 28°, 53°, 73°, 90°, 112°, 135°, 157°, 180°, 208°, 233°, 253°, 270°, 292°, 315° and 337°, with an achromatic adaptation point (Illuminant C). On this circle, the L-, M- and S-cone contrasts were sinusoidal functions of chromaticity with maximum attainable values of 8%, 16% and 83%, respectively.

Cell selection. Only cells with consistent receptive fields in successive stimulation blocks were included in the data set. All included cells had receptive fields with a signal-to-noise (S/N) ratio of at least +22 dB. The S/N ratio was calculated as the ratio of the power (sum of squares) of the receptive field at the optimal delay (signal) to the power of the receptive field at zero delay (noise). Nearly all cells in the data set gave chromatic responses that were equally strong or stronger than their responses to a spatially uniform, 20% contrast, achromatic stimulus.

Chromatic-temporal receptive-field visualization. Chromatic-temporal receptive-field maps were computed on a 16 × 40 grid and subsequently interpolated and smoothed. Magnitude of firing probability is colour-coded: grey corresponds to random, uncorrelated firing; warm colours (red to yellow to white) correspond to progressively higher-than-random firing probability; cool colours (light blue to dark blue to black) correspond to progressively lower-than-random firing probability. Iso-response contours are drawn at intervals of 6.25% correlation level.

Curve fitting. To extract times-to-peak (in separate receptive field slices), receptive-field slices parallel to the time axis (τ) were fit with a function that describes a damped oscillation, $F(\tau): F(\tau) = A \times \text{Env}(\tau, \tau_p; \tau_r; \tau_d) \times \sin(2\pi \times (f \times \tau - \xi/360))$, with

$$\text{Env}(\tau, \tau_p; \tau_r; \tau_d) = \begin{cases} e^{-0.5 \left(\frac{\tau - \tau_p}{\tau_r} \right)^2}, & \tau \leq \tau_p \\ e^{-0.5 \left(\frac{\tau - \tau_p}{\tau_d} \right)^2}, & \tau > \tau_p \end{cases}$$

where A is the response gain, τ_p is the envelope's peak latency, τ_r and τ_d are time constants for the rising and decaying phases of the damping envelope, respectively, and f and ξ are the temporal frequency and temporal phase of the oscillation, respectively. The slice's time-to-peak was extracted by searching for the τ value for which $F(\tau)$ reached its maximum. Separate receptive-field slice fitting was done to extract time-to-peak values for all chromatic temporal receptive fields, including those whose unusual structure (90° chromatic shifts) precluded fitting of the entire receptive field, and to extract a more complete set of time-to-peak values for all the different chromaticities tested. For single-peaked chromatic-temporal receptive fields (Fig. 1A, B) three slices were fit, corresponding to the three chromaticities of maximum response. For double-peaked receptive fields (Fig. 1C), six slices were fit, three around each chromatic peak. For cells with continuously changing responses latencies (Fig. 1D), three slices were fit around their first strong peak.

To extract receptive-field peak chromaticity, receptive-field time-to-peak, and response exponent (entire receptive-field fit), receptive-field slices parallel to the chromaticity axis (ϕ) were taken at successive time intervals (τ_i) throughout the cell's entire response duration and fit with a function, $G(\phi, \tau_i)$, that describes an exponentiated sinusoid with a threshold: $G(\phi, \tau_i) = A(\tau_i) \times \text{Max}\{\text{Sgn}(f(\phi, \tau_i)) \times |f(\phi, \tau_i)|^{n(\tau_i)}, T\}$, $\tau_i = 0, 4, \dots, 220$ ms, with

$$f(\phi, \tau_i) = \sin(2\pi \times (\phi - \phi(\tau_i))/360), \quad \text{Sgn}(x) = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

where $A(\tau_i)$, $n(\tau_i)$ and $\phi(\tau_i)$ are the response gain, the response exponent, and the chromatic axis of the receptive field, respectively, at a stimulus-response delay of τ_i ms, and T is the response threshold ($T = 0$ for half-wave rectified responses). For neurons with relatively constant chromatic preference, receptive-field peak chromaticity was obtained by averaging $\phi(\tau_i)$ over a time window in which $A(\tau_i)$ was within 80% of its maximum value. Receptive-field time-to-peak was obtained by finding the τ_i value for which $A(\tau_i)$ reached its maximum.

Received 22 June; accepted 20 August 1998.

- De Valois, R. L. *et al.* Analysis of response patterns of LGN cells. *J. Opt. Soc. Am.* **56**, 966–977 (1966).
- Lennie, P. & D'Zmura, M. Mechanisms of color vision. *CRC Crit. Rev. Neurobiol.* **3**, 333–400 (1988).
- Dacey, D. Circuitry for color coding in the primate retina. *Proc. Natl Acad. Sci. USA* **93**, 582–585 (1996).

- Lennie, P. *et al.* Chromatic mechanisms in striate cortex of macaque. *J. Neurosci.* **10**, 649–669 (1990).
- De Valois, R. L. & De Valois, K. K. A multi stage color model. *Vision Res.* **33**, 1053–1065 (1993).
- Hartline, H. K. The receptive fields of optic nerve fibers. *Am. J. Physiol.* **130**, 690–699 (1940).
- MacLeod, D. I. A. & Boynton, R. M. Chromaticity diagram showing cone excitation by stimuli of equal luminance. *J. Opt. Soc. Am.* **69**, 1183–1186 (1979).
- Derrington, A. M. *et al.* Chromatic mechanisms in the lateral geniculate nucleus of macaque. *J. Physiol. (Lond.)* **357**, 241–265 (1984).
- de Boer, E. & Kuypers, P. Triggered correlation. *IEEE Trans. Biomed. Eng.* **15**, 169–179 (1968).
- Ringach, D. L. *et al.* Dynamics of orientation tuning in macaque primary visual cortex. *Nature* **387**, 281–284 (1997).
- Reid, R. C. *et al.* The use of m-sequences in the analysis of visual neurons: linear receptive field properties. *Vis. Neurosci.* **14**, 1015–1027 (1997).
- Stockman, A. *et al.* The temporal properties of the human short-wave photoreceptors and their associated pathways. *Vision Res.* **31**, 189–208 (1991).
- Gegenfurtner, K. R. & Hawken, M. J. Temporal and chromatic properties of motion mechanisms. *Vision Res.* **35**, 1547–1563 (1995).
- Rabin, J. *et al.* Visual evoked potentials in three-dimensional color space: correlates of spatio-chromatic processing. *Vision Res.* **34**, 2657–2671 (1994).
- Dacey, D. M. & Lee, B. B. The 'blue-on' opponent pathways in primate retina originates from a distinct bistratified ganglion cell. *Nature* **367**, 731–735 (1994).
- Martin, P. R. *et al.* Evidence that blue-on cells are part of the third geniculocortical pathway in primates. *Eur. J. Neurosci.* **9**, 1536–1541 (1997).
- Gielen, C. C. A. M. *et al.* Reconstruction of cone-system contributions to responses of colour-opponent neurons in monkey lateral geniculate. *Biol. Cybern.* **44**, 211–221 (1982).
- Tolhurst, D. J. & Heeger, D. J. Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex. *Vis. Neurosci.* **14**, 293–309 (1997).
- Albrecht, D. G. & Geisler, W. S. Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Vision Res.* **7**, 531–546 (1991).
- McCormick, D. A. *et al.* Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J. Neurophys.* **54**, 782–806 (1985).
- Pugh, M. *et al.* Computational modeling of orientation tuning dynamics in V1 neurons. *Soc. Neurosci. Abstr.* **23**, 603 (1997).
- de Monasterio, F. M. *et al.* Staining of blue-sensitive cones of the macaque retina by a fluorescent dye. *Science* **213**, 1278–1281 (1981).
- Calins, D. J. *et al.* Microcircuitry and mosaic of a blue-yellow ganglion cell in the primate retina. *J. Neurosci.* **18**, 3373–3385 (1998).
- Malpe, J. G. & Schiller, P. H. Lack of blue off-center cells in the visual system of the monkey. *Brain Res.* **141**, 385–389 (1978).
- Lennie, P. Recent developments in the physiology of color vision. *Trends Neurosci.* **5**, 243–248 (1984).
- De Valois, R. L. *et al.* Color appearance with and without S-opponent cells. *Vision Res.* (submitted).
- Hawken, M. J. *et al.* Laminar organization and contrast sensitivity of direction-selective cells in the striate cortex of the old world monkey. *J. Neurosci.* **8**, 3541–3548 (1988).

Acknowledgements. We thank K. K. De Valois, S. D. Elfar, J. Gallant and E. Switkes for comments on the manuscript, and E. Switkes for implementation of the MBDKL colour space. This work was funded by the NIH.

Correspondence and requests for materials should be addressed to N.P.C. (e-mail: nicolas@valois.berkeley.edu).

Selective activation of Ca^{2+} -activated K^+ channels by co-localized Ca^{2+} channels in hippocampal neurons

Neil V. Marrion & Steven J. Tavalin

Vollum Institute, Oregon Health Sciences University Portland, Oregon 97201, USA

Department of Pharmacology, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, UK

Calcium entry through voltage-gated calcium channels can activate either large- (BK) or small- (SK) conductance calcium-activated potassium channels. In hippocampal neurons, activation of BK channels underlies the falling phase of an action potential and generation of the fast afterhyperpolarization (AHP)^{1,2}. In contrast, SK channel activation underlies generation of the slow AHP after a burst of action potentials³. The source of calcium for BK channel activation is unknown, but the slow AHP is blocked by dihydropyridine antagonists^{4,5}, indicating that L-type calcium channels provide the calcium for activation of SK channels. It is not understood how this specialized coupling between calcium and potassium channels is achieved. Here we study channel activity in cell-attached patches from hippocampal neurons and report a unique specificity of coupling. L-type channels activate SK channels only, without activating BK chan-

nels present in the same patch. The delay between the opening of L-type channels and SK channels indicates that these channels are 50–150 nm apart. In contrast, N-type calcium channels activate BK channels only, with opening of the two channel types being nearly coincident. This temporal association indicates that N and BK channels are very close. Finally, P/Q-type calcium channels do not couple to either SK or BK channels. These data indicate an absolute segregation of coupling between channels, and illustrate the functional importance of submembrane calcium microdomains.

We studied the functional coupling of somatic inward calcium channels to outward calcium-activated potassium channels in acutely isolated CA1 hippocampal pyramidal neurons, to determine how specific calcium-channel subtypes selectively activate BK or SK channels. L-type calcium channels were activated by depolarization and visualized, using 160 or 10 mM Ca^{2+} in the pipette solution (267 of 502 patches). Evoked channels had an amplitude of -216 ± 40 fA at 0 mM in 160 mM Ca^{2+} (mean $\pm$ s.d., $n = 9$ patches; events >1 ms duration⁶) and were of extremely short duration (mean duration 439 μs , see Methods) (Fig. 1a). In approximately 50% of the patches, outward channel openings followed the opening of an L-type channel (125 out of 267 patches) (Fig. 1a, d). The outward channel opening followed a resolved L-type channel opening in $>99\%$ of the sweeps. The outward channel amplitude was 385 ± 59 fA (mean $\pm$ s.d., $n = 28$ patches) at 0 mV. The single-channel current–voltage relationship yielded a limiting slope conductance of 10 pS (0 to +40 mV), consistent with that of both cloned SK channels^{7,8} and hippocampal CA1 SK channels (B. Hirschberg and N.V.M., unpublished observations). Long duration L-type openings induced by BAY K8644 also activated SK channels, supporting the proposal that L-type channels provide the calcium for activation of SK channels (Fig. 1c). Subsequent bath application of nimodipine (3 μM) inhibited both calcium and potassium channel activity, indicating that SK channel opening required a preceding L-type channel opening ($n = 8$) (Fig. 1b). Coupling between L-type calcium channels and SK channels was also observed using a more physiological concentration of Ca^{2+} (10 mM) (30 out of 49 patches) (Fig. 1d).

After we observed coupling in the cell-attached configuration (Fig. 2a), we excised inside-out patches to demonstrate the calcium dependence of coupled SK channels. In the presence of 100 nM Ca^{2+} , SK channel activity of short duration was rarely observed (Fig. 2b, sweep 1). Both SK and BK channels were activated markedly by application of 600 nM Ca^{2+} (Fig. 2b, sweeps 2–5). SK channels exhibited bursts of openings of long duration (Fig. 2b, sweeps 2–4), with activity being independent of voltage (data not shown). These properties, including insensitivity to the potassium channel blockers used (see Methods), indicate that these channels are SK channels. BK channel activity was also observed after excision, although no coupling to L-type channel activation was observed in the cell-attached configuration (Fig. 2b, sweep 5). Of 114 patches with coupled L-type and SK channels, 56 were also found to contain BK channels that were activated either by stepping the membrane voltage to +100 mV (200 ms duration) or by patch excision (Fig. 2). In none of these patches were BK channel openings coupled to L-type channel openings.

A delay often occurred between the closure of an L-type channel and the opening of an SK channel (Figs 1 and 3Aa). Analysis of this delay provides an estimate of the distance between the two channels (Fig. 3A). The first latency distribution to SK channel opening was best fit by the sum of three exponentials (Fig. 3Ab), with time constants (τ) identical to those observed for the closed time distribution of cloned SK channels exposed to 1 μM Ca^{2+} (ref. 8). This indicates that $\sim 1 \mu\text{M}$ Ca^{2+} was available at the SK channel from the 'priming' L-type channel opening. These data also indicate that the activation kinetics of the SK channel dictate the delay between the two events, the observed latency to SK channel opening being consistent with the activation time constant for cloned SK

channels in 1 μM Ca^{2+} (ref. 8). An estimated calcium concentration of 1 μM at the SK channel would produce an open probability (P_o) of 0.4–0.6 (ref. 9 and B. Hirschberg and N.V.M., unpublished observations), a P_o estimated to occur at the peak of the slow AHP^{10,11}. Models of calcium microdomains predict that a Ca^{2+} concentration of $\sim 1 \mu\text{M}$ is reached at a distance of 50–150 nm from the pore of the calcium channel^{12,13}, indicating that this distance probably separates L-type and SK channels.

If L-type and SK channels are 50–150 nm apart, channel coupling should be disrupted by the intracellular addition of the rapid, high-affinity Ca^{2+} buffer BAPTA. Preincubation of cells with the

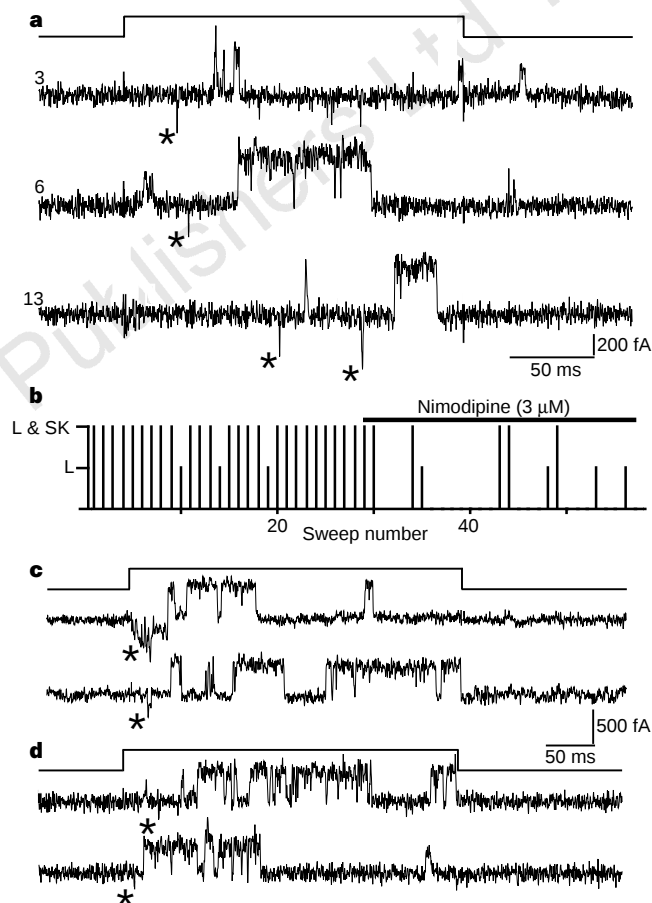


Figure 1 Selective activation of SK channels by co-localized L-type calcium channels in cell-attached patches. **a**, Upon a step depolarization to 0 mV from a holding potential of -40 mV, brief inward channel openings (labelled with asterisks) preceded activation of a small outward channel. Numbers indicate sweep number (see **b**). **b**, Diary plot of channel activity during the recording shown in **a**. Openings of L-type channels are represented as shorter lines, whereas sweeps exhibiting both L and SK channel openings are shown as longer. Before addition of nimodipine, all sweeps exhibited L-type channel activity, with most coupled to SK channel activation. Upon addition of nimodipine (3 μM), the number of sweeps exhibiting both inward and outward channel activity was greatly reduced, indicating that the activity of the outward channel depended on the prior opening of a preceding L-type channel. **c**, Application of BAY K8644 (1 μM) induced long duration inward channel openings (labelled with asterisks) characteristic of modified L-type calcium channels. Long duration openings coupled to SK channel activation, with a delay still apparent. The openings shown represent the longer openings observed; most openings showed a more modest prolongation of open time²⁴. **d**, Coupling of L-type channels (labelled with asterisks) to SK channels observed using a more physiological calcium concentration (10 mM Ca^{2+} in the pipette solution). Inward channel openings were of smaller amplitude than with 160 mM Ca^{2+} in the pipette (**a**); however, the openings seen here were still capable of activating SK channels within the patch. In **c**, **d**, channel activity was evoked by depolarization to 0 mV from -60 mV.

membrane-permeant BAPTA-AM (10 μ M) for 30 min increased the percentage of patches exhibiting L-type channels from 50% to 100% (22 out of 22). The proportion of patches exhibiting both L-type and SK channels remained near 50% (10 out of 22), although coupling was disrupted over the time of recording (10 out of 10 patches). Figure 3Ba shows diary plots taken from two patches, one from a cell preincubated with BAPTA-AM and the other from a control cell. The most marked effect of intracellular BAPTA was to enhance dramatically the number of sweeps that showed L-type openings. In the first minute of recording an enhanced coupling to SK channels was observed. This was unexpected, and was probably the result of calcium-bound BAPTA behaving as a mobile buffer to enhance the spatial diffusion of calcium^{13,14}. However, during the remainder of the recording, coupling of L-type and SK channels was inhibited. This was observed as fewer sweeps containing SK channel

activity (Fig. 3Ba, b), which probably reflects a slow re-equilibration of free BAPTA into the patch, restoring buffering of Ca^{2+} . In contrast, the extent of L-type and SK channel coupling remained intact during the control patch recording (Fig. 3Ba). Examination of individual sweeps from the BAPTA-treated patch showed that inward L-type channel openings became prolonged as coupled SK channel activity subsided (Fig. 3Bb). Finally, the duration of the L-type openings became very long, whereas SK channel activity was almost lost (Fig. 3Bb). These data confirm that L-type and SK channels are sufficiently separated to allow disruption of coupling by intracellular BAPTA.

N-type channels were resolved infrequently (16 out of 61 patches) upon depolarization from negative holding potentials (-60 to -100 mV), consistent with their limited somatic expression in hippocampal neurons^{15,16}. These channels exhibited a single-channel amplitude of -424 ± 51 fA at 0 mV in 160 mM Ca^{2+}

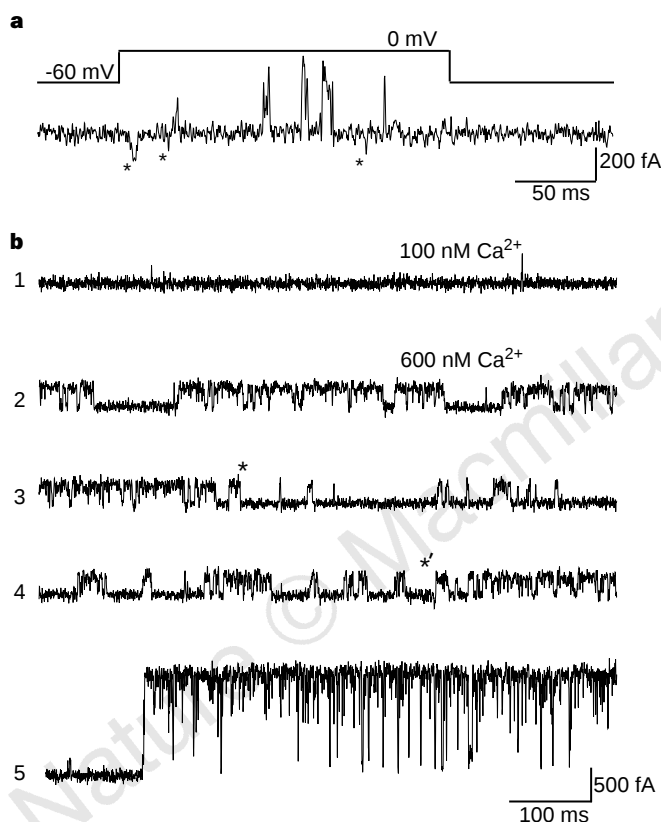


Figure 2 SK and BK channels are activated in a patch after excision. **a**, After observing coupling between L-type and SK channels (marked by asterisks) in the cell-attached-patch configuration, the patch was excised to the inside-out configuration (**b**). **b**, At 0 mV and in the presence of 100 nM Ca^{2+} , short duration outward channel openings of ~ 400 fA in amplitude were rarely observed (sweep 1). Application of 600 nM Ca^{2+} to the inside-out patch caused activation of both an SK channel (sweeps 2 and 3) and a BK channel (sweep 5). SK channel activity consisted of long duration bursts of openings, but showed non-stationary gating similar to that shown by cloned SK channels⁸. The channel was observed to enter a low P_o short open-time activity (entrance into behaviour shown in sweep 3, marked with an asterisk). Exit from this behaviour was observed several seconds later as long bursts of activity (sweep 4, marked with *). A BK channel was activated by application of 600 nM Ca^{2+} to the inside-out patch (sweep 5), although no BK channel openings were observed in the cell-attached configuration.

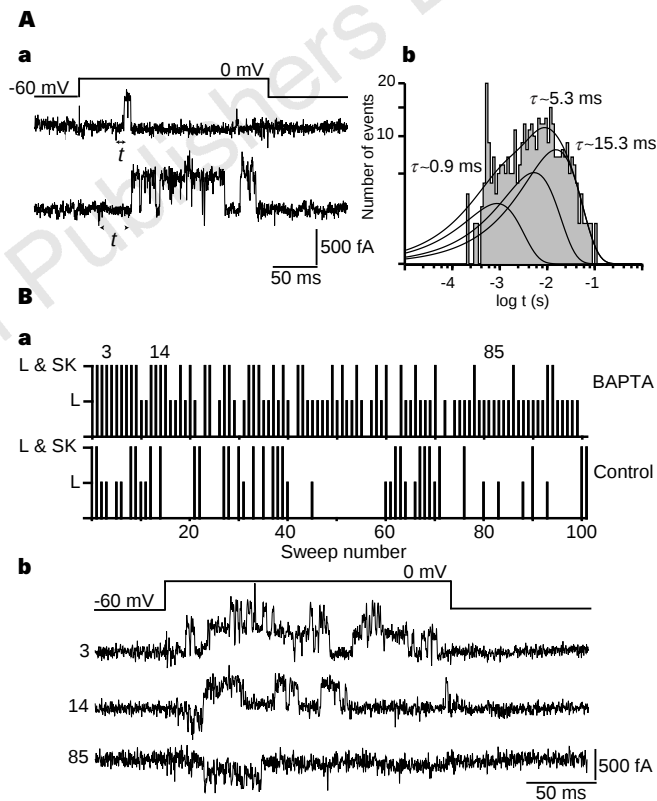


Figure 3 L-type and SK channels are separated by 50–150 nm. **Aa**, Examples of coupled L-type and SK channels evoked by voltage steps from -60 mV to 0 mV. The time (t) measured between the closing of the last L-type channel and the opening of the first SK channel is shown. **b**, the histogram of latencies between channel events was best fit by the sum of three exponentials with time constants (τ) 0.9, 5.3 and 15.3 ms. The number of statistically significant exponentials was determined using the method of maximum likelihood ratios²⁶. **B**, Coupling between L-type and SK channels was disrupted by intracellular BAPTA. **a**, Diary plots of the BAPTA-treated patch shown in **b** and a control patch. Activity of L-type channels was greatly increased in cell-attached patches from BAPTA-treated neurons. With time, the coupling to SK channels was inhibited. In contrast, the coupling remained intact throughout the recording in patches from control cells. **b**, Examples of activity evoked by depolarization from -60 mV to 0 mV from the patch shown in **a**. During the first sweeps, an enhanced activation of SK channels was observed; activation declined as the duration of L-type openings increased. Numbers indicate sweep number (see **a**). The data in **B** show that L-type channel openings experience a high degree of Ca^{2+} -dependent inactivation. The prolongation of channel openings during recording shows that calcium-dependent inactivation normally determines the apparent duration of opening²⁶. The increase in the number of patches exhibiting L-type channels during BAPTA treatment probably results from the L-type channel open time being prolonged, allowing improved resolution of the brief inward channel opening.

(mean $\pm$ s.d., $n = 8$ patches; events >1 ms duration⁶) and were of short duration (mean duration 677 μ s; see Methods). In a small proportion of patches containing N-type channels (4 out of 16), BK channel activation was nearly coincident with inward channel opening (Fig. 4A), a temporal association that indicates that these channels were extremely close to each other. This was confirmed by the observation that intracellular BAPTA did not disrupt coupling between the two channel types (in 9 out of 28 patches from BAPTA-

treated cells in which N-type channels were present, N-BK coupling was observed) (Fig. 4B). This suggests that N-type and BK channels are separated by a distance shorter than the buffering length constant for BAPTA (~ 30 nm under these conditions¹³). Therefore, we suggest that coupled N-type and BK channels may be located within 30 nm of each other.

Selective coupling of L-SK and N-BK channels was observed within a single patch (Fig. 4C). BK channel activation was coincident

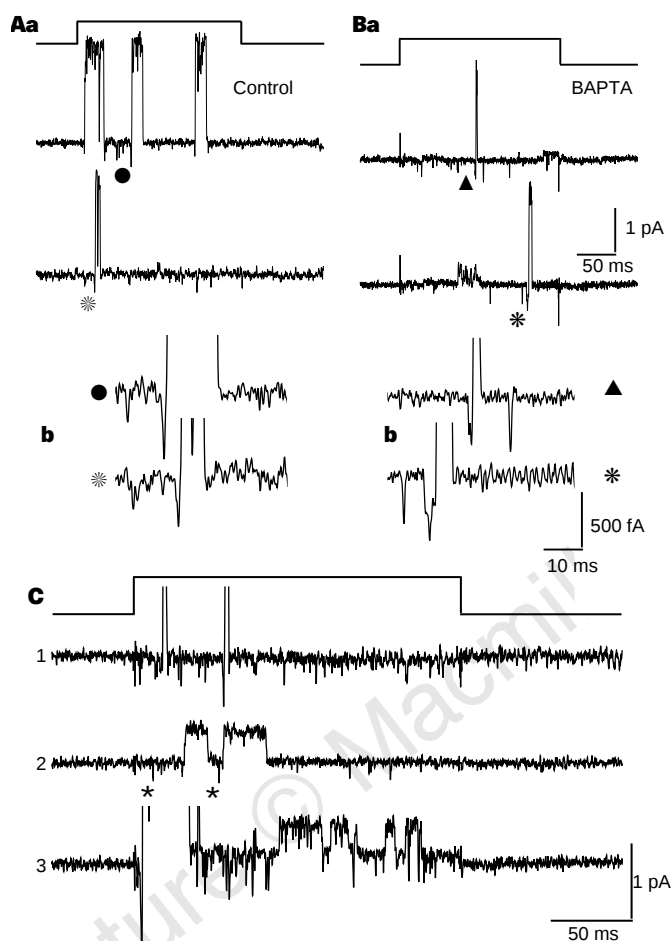


Figure 4 N-type channels activate BK channels. **Aa**, Step depolarization from -60 to 0 mV revealed BK channel activation coincident with the opening of inward N-type calcium channels. The close temporal association of these two channels can be seen in the expanded traces shown in **b**, in which the full amplitude of the BK channel opening has been truncated to allow closer study of the coincident openings. This near coincident opening of N-type and BK channels indicates that they may be very close to each other. **Ba**, Intracellular BAPTA does not disrupt the coupling between N-type and BK channels. Current sweeps recorded from a cell-attached patch of a BAPTA-AM-treated neuron are shown. Depolarization to 0 mV from a holding potential of -100 mV revealed N-type-channel openings with near coincident openings of a BK channel. The close temporal coupling between the two channel types can be seen in the expanded traces shown in **b**. The inability of intracellular BAPTA to disrupt the coupling of N-type to BK channels confirms that these two channel types are close to each other. **C**, Patch current sweeps obtained from a cell not treated with nimodipine or ω -conotoxin GVIA. Depolarization to 0 mV from -60 mV evoked both N-type and L-type channel activity that coupled to separate calcium-activated potassium channels. Sweep 1 shows N-type channel openings, with BK channel openings appearing nearly coincident. The amplitude of the BK channel has been truncated for clearer illustration of the temporal coupling between the N-type and BK channels. Sweep 2 shows only L-type channel openings (labelled with asterisks) coupling to the activation of an SK channel only. Sweep 3 shows N-type channel openings with coincident BK channel activation, and SK channel activation which is presumably the result of a preceding L-type channel opening.

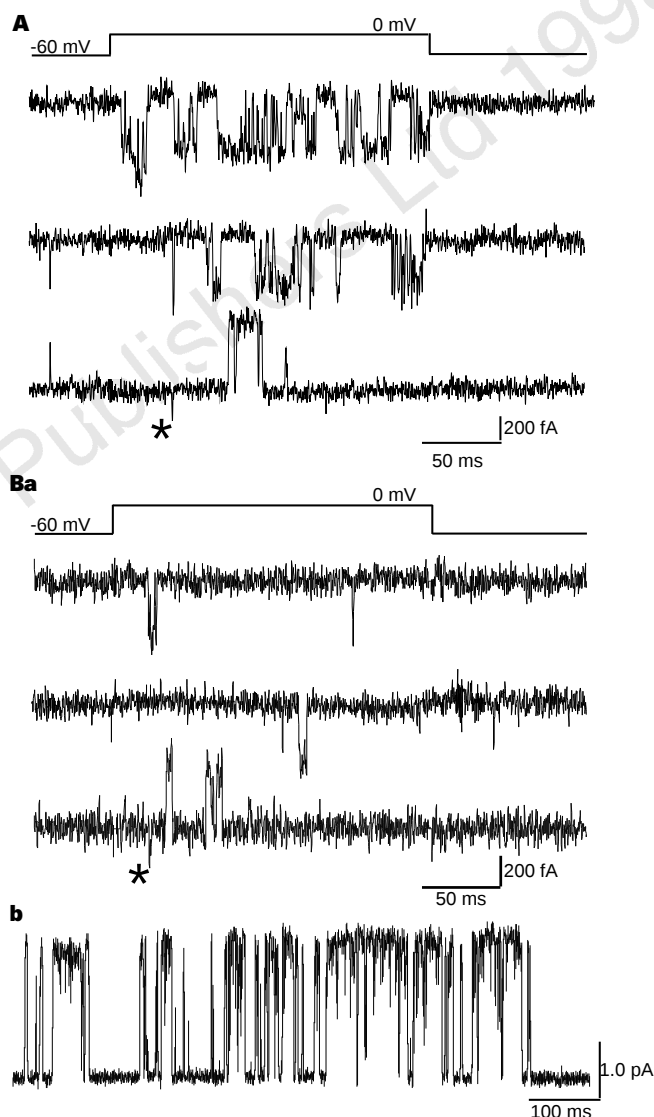


Figure 5 P/Q-type calcium channels do not activate calcium-activated potassium channels. **A**, P/Q-type channel openings were evoked by step depolarization from -60 mV to 0 mV. Openings were of both longer duration and greater amplitude than L-type channel openings (see bottom trace). However, despite these properties, P/Q-type channels did not activate an SK channel that was shown to be present in the patch by the opening of a single L-type channel (marked with an asterisk) (bottom trace). **Ba**, The opening of a P/Q-type channel could not activate either an SK channel or a BK channel present in the cell-attached patch. SK channel activity was seen after a rare opening of a coupled L-type channel (marked with an asterisk). **b**, After the patch was excised to the inside-out configuration, BK channel activity was observed at 0 mV in the presence of 1μ M Ca^{2+} .

with N-type channel openings (Fig. 4C, sweeps 1, 3). In a limited number of sweeps, L-type channel openings were observed to couple to SK channel activation (Fig. 4C, sweep 2). In one sweep, both N-BK and L-SK coupling was observed (Fig. 4C, sweep 3).

P/Q-type channels were rarely observed, as expected from the limited somatic labelling by antibodies to these channels¹⁷. In 9 out of 68 patches, voltage jumps from -60 mV to 0 mV evoked inward channel currents with an amplitude of -443 ± 10 fA in 160 mM Ca^{2+} (mean $\pm$ s.d., $n = 4$ patches; events >1 ms duration⁶) and a mean duration of 2.32 ms (see Methods) (Fig. 5). In all examples, P/Q channel openings failed to activate either SK or BK channels. Figure 5A shows an example of a P/Q-type channel that was ineffective at activating an SK channel, with the presence of an SK channel being confirmed by a rare opening of a coupled L-type channel. Figure 5B shows a patch in which opening of a P/Q-type channel was unable to activate either an SK or a BK channel. SK channel activation was seen following a rare opening of an L-type channel (Fig. 5Ba), and BK channel activity was observed only after patch excision (Fig. 5Bb). In seven more patches in which P/Q openings were observed, the presence of outward channels was confirmed; SK channels were identified upon patch excision or by coupling to an L-type opening, while BK channels were found to be present either by patch excision or after voltage steps to $+100$ mV. Therefore, despite showing a larger unitary current and longer bursts of openings than L- or N-type channels, somatic P/Q-type channels do not activate either SK or BK channels.

The proportion of patches containing BK channels (46%, 103 out of 224 tested) was much greater than that containing N-type calcium channels (26%, 16 out of 61). In addition, many patches containing N-type channels were not observed to couple to BK channels (see above). This indicates that many N-type and BK channels are not in close association with each other. Also, $\sim 50\%$ of patches displayed L-type channel activity without associated SK channel openings, suggesting that there are 'spare' L-type channels that are not associated with SK channels. In contrast, the number of patches shown to contain SK channels upon patch excision was similar to the proportion of SK channels coupled to L-type channels (B. Hirschberg and N.V.M., unpublished observations). This indicates that most SK channels are functionally coupled to L-type channels, with few being unassociated.

The proximity of N-type and BK channels allows BK channels to be activated rapidly by the opening of an N-type calcium channel. This is in agreement with the physiological role of BK channels, allowing them to be activated during the falling phase of the action potential to cause repolarization of the soma membrane potential and generation of the fast AHP. In contrast, we propose that L-type and SK channels are 50 – 150 nm apart, but this seems inconsistent with the slow rise time of the slow AHP^{3,9,10}. In addition, this afterpotential is observed at negative potentials, at which L-type channels are not normally expected to be active. However, trains of action potential waveforms induce delayed facilitation of L-type calcium channels, which is characterized by prolonged L-type channel activity at membrane potentials negative to -60 mV (ref. 6). The time course of delayed facilitation and the slow AHP are the same, with both exhibiting a slow rising phase and decay^{3,6,9,10}. Therefore, it is possible that delayed facilitation dictates the time course of SK channel activation, producing the characteristic slow rise and decay of the slow AHP.

We have shown the selective coupling of calcium channel subtypes to calcium-activated potassium channels. Coupling of calcium channel subtypes and calcium-activated potassium channels may differ between cell types^{18,19}. Segregation of coupling between channel types is proposed to result from microdomains of submembrane calcium. The presence of microdomains has been proposed¹² and may provide inhibitory coupling between L-type calcium channels in the heart²⁰. We have provided the first evidence, to our knowledge, of channel localization using submembrane

calcium microdomains to achieve selective functional coupling between different types of channels. □

Methods

Acutely dissociated hippocampal CA1 neurons were prepared as described⁶. Calcium channel subtypes were isolated using a combination of channel blockers. To isolate L-type channels, N- and P/Q-type channels were blocked by preincubation of cells in dissociation solution⁶ supplemented with 0.1 mg ml⁻¹ BSA and 10 mM D-glucose, with ω -conotoxins GVIA (1 μ M)²¹ and MVIIC (5 μ M)²², for 30 min. The dissociation solution contained low divalent ion concentrations⁶, to prevent these ions antagonizing the channel block²². Similarly, the pipette solution was supplemented with ω -conotoxin GVIA (5 μ M) to prevent antagonism of block by the high concentration of Ca^{2+} used in the electrode. To study N-type channels in isolation, L- and P/Q-type channels were blocked by preincubation with nimodipine (3 μ M) and ω -conotoxin MVIIC (5 μ M) for 30 min (see above). Block of P/Q-type channels is prolonged ($\tau_{\text{off}} \sim 200$ min)²², whereas block of N-type channels by ω -conotoxin MVIIC is reversible ($\tau_{\text{off}} \sim 30$ s). Therefore, washing the cells with toxin-free external solution caused dissociation of ω -conotoxin MVIIC from N-type channels, leaving P/Q-type channels blocked. Cells were used for only 1 h after the preincubation to prevent contamination of N-type-channel recording with unblocking P/Q-type channels. L-type-channel block was continued by including nimodipine (3 μ M) in the pipette and external solutions. Finally, to isolate P/Q-type channels, L- and N-type channels were blocked by preincubation with nimodipine (3 μ M) and ω -conotoxin GVIA (1 μ M) for 30 min. Nimodipine (3 μ M) and ω -conotoxin GVIA (5 μ M) were included in the pipette solution, and nimodipine (3 μ M) was also added to the external medium.

Voltage-dependent potassium channels were blocked by the addition of α - and β -dendrotoxin (200 nM) to the preincubation and pipette solutions. In addition, external and pipette solutions contained 4 -aminopyridine (1 mM) and $3,4$ -diaminopyridine (1 mM). Finally, calcium entry through voltage-dependent sodium channels was prevented by including tetrodotoxin (10 μ M) in the pipette solution. Pharmacological isolation of each calcium channel subtype allowed the identification of distinct biophysical properties. L-type channels possessed a slope conductance of ~ 2.5 pS (in 160 mM Ca^{2+}), and a bi-exponential open-time distribution with time constants 0.49 and 4.0 ms (ref. 6). In contrast, N-type channels possessed a slope conductance of 8 pS, a mono-exponential open-time distribution ($\tau = 0.52$ ms) and steady-state inactivation at positive holding potentials (-40 to -20 mV). Finally, P/Q-type channels exhibited a slope conductance of 8 – 10 pS, a mono-exponential open-time distribution ($\tau = 2.32$ ms) and characteristic bursts of opening (Fig. 5). BAPTA-AM was dissolved in dimethyl sulphoxide (DMSO) at a concentration of 10 mM. Cells were treated with the acetoxymethylester by including it in the preincubation at 10 μ M for 30 min.

Cells were washed and superfused (15 ml min⁻¹) with an external solution containing (in mM): potassium methyl sulphate, 125 ; KCl, 35 ; MgCl_2 , 5 ; Na-HEPES, 10 ; EGTA, 10 ; CaCl_2 , 5.64 (to give an estimated free concentration of 60 nM) (pH 7.4 with KOH). Cells in this solution had a membrane potential of ~ 0 mV. All potentials are expressed as the negative of the potential imposed on the pipette, without correction for a -17 mV (for 160 mM Ca^{2+}) or a $+16$ mV (for 10 mM Ca^{2+}) liquid junction potential. Membrane patches were first excised into a Mg^{2+} -free superfusion solution containing 100 nM Ca^{2+} . Once the patch had stabilized, the free Ca^{2+} concentration was raised to 600 or $1,000$ nM.

Cell-attached patch recordings were made using thick walled (outer diameter 1.5 mm, inner diameter 0.5 mm) quartz electrodes (7 – 10 M Ω) containing (in mM): CaCl_2 , 160 or 10 ; KOH, 0.5 or 2 ; N-methyl D-glutamine, 150 or 0 ; Na-HEPES, 10 ; 4 -aminopyridine, 1 ; $3,4$ -diaminopyridine, 1 ; α -dendrotoxin, 200 nM; β -dendrotoxin, 200 nM and tetrodotoxin, 10 μ M (pH 7.4). Single-channel currents were recorded with an Axopatch 200A amplifier (Axon Instruments), filtered at 1 – 2 kHz with an 8 -pole Bessel filter (frequency Devices, MA) and acquired at intervals for analysis using Pulse (Heka, distributed by Instrutech Corp.) onto a Quadra 800 (Apple Computer). Single channels were analysed using MacTAC (Bruxton Corp., distributed by Instrutech)²³. The intersweep interval was 5 s in all cases.

All reagents were obtained from Sigma, except ω -conotoxin MVIIC (Bachem), ω -conotoxin GVIA (Peninsula), α - and β -dendrotoxin (Alomone), CaCl_2 (Fluka), nimodipine, BAPTA-AM, BAY K8644 and HEPES (Calbiochem).

Received 3 June; accepted 25 August 1998.

- Storm, J. F. Action potential repolarization and a fast after-hyperpolarization in rat hippocampal pyramidal cells. *J. Physiol. (Lond.)* **385**, 733–759 (1987).
- Lancaster, B. & Nicoll, R. A. Properties of two calcium-activated hyperpolarizations in rat hippocampal neurones. *J. Physiol. (Lond.)* **389**, 187–203 (1987).
- Lancaster, B. & Adams, P. R. Calcium-dependent current generating the afterhyperpolarization of hippocampal neurones. *J. Neurophysiol.* **55**, 1268–1282 (1986).
- Moyer, J. R., Thompson, L. T., Black, J. P. & Disterhoft, J. F. Nimodipine increases excitability of rabbit CA1 pyramidal neurons in an age- and concentration-dependent manner. *J. Neurophysiol.* **68**, 2100–2109 (1992).
- Rascol, O., Potier, B., Lamour, Y. & Dutar, P. Effects of calcium agonist and antagonists on calcium-dependent events in CA1 hippocampal neurons. *Fundamen. Clin. Pharmacol.* **5**, 299–317 (1991).
- Cloues, R. K., Tavalin, S. J. & Marrion, N. V. β -Adrenergic stimulation selectively inhibits long-lasting L-type calcium channel facilitation in hippocampal pyramidal neurons. *J. Neurosci.* **17**, 6493–6503 (1997).
- Köhler, M. et al. Small conductance, calcium-activated potassium channels from mammalian brain. *Science* **273**, 1709–1714 (1996).
- Hirschberg, B., Maylie, J., Adelman, J. P. & Marrion, N. V. Gating of recombinant small-conductance Ca-activated K^+ channels by calcium. *J. Gen. Physiol.* **111**, 565–581 (1998).
- Lancaster, B., Nicoll, R. A. & Perkel, D. J. Calcium activates two types of potassium channels in rat hippocampal neurons in culture. *J. Neurosci.* **11**, 23–30 (1991).
- Sah, P. & Isaacson, J. S. Channels underlying the slow afterhyperpolarization in hippocampal pyramidal neurons: neurotransmitters modulate the open probability. *Neuron* **15**, 435–441 (1995).
- Valiante, T. A., Abdul-Ghani, M. A., Carlen, P. L. & Pennefather, P. Analysis of current fluctuations during after-hyperpolarization current in dentate granule neurones of the rat hippocampus. *J. Physiol. (Lond.)* **499**, 121–134 (1997).
- Chad, J. E. & Eckert, R. Calcium domains associated with individual channels can account for anomalous voltage relations of Ca-dependent responses. *Biophys. J.* **45**, 993–999 (1984).
- Naraghi, M. & Neher, E. Linearized buffered Ca^{2+} diffusion in microdomains and its implications for calculation of $[Ca^{2+}]$ at the mouth of a calcium channel. *J. Neurosci.* **17**, 6961–6973 (1997).
- Zhang, L. et al. Potentiation of a slow Ca^{2+} -dependent K^+ current by intracellular Ca^{2+} chelators in hippocampal CA1 neurons of rat brain slices. *J. Neurophys.* **74**, 2225–2241 (1995).
- Westenbroek, R. E. et al. Biochemical properties and subcellular distribution of an N-type calcium channel $\alpha 1$ subunit. *Neuron* **9**, 1099–1115 (1992).
- Mills, L. R. et al. N-type Ca^{2+} channels are located on somata, dendrites, and a subpopulation of dendritic spines on live hippocampal pyramidal neurons. *J. Neurosci.* **14**, 6815–6824 (1994).
- Elliott, E. M., Malouf, A. T. & Catterall, W. A. Role of calcium channel subtypes in calcium transients in hippocampal CA3 neurons. *J. Neurosci.* **15**, 6433–6444 (1995).
- Davies, P. J., Ireland, D. R. & McLachlan, E. M. Sources of Ca^{2+} for different Ca^{2+} -activated K^+ conductances in neurones of the rat superior cervical ganglion. *J. Physiol. (Lond.)* **495**, 366–386 (1996).
- Pineda, J. C., Waters, R. S. & Foehring, R. C. Specificity in the interaction of HVA Ca^{2+} channel types with Ca^{2+} -dependent AHPs and firing behavior in neocortical pyramidal neurons. *J. Neurophysiol.* **79**, 2522–2534 (1998).
- Imredy, J. P. & Yue, D. T. Submicroscopic Ca^{2+} diffusion mediates inhibitory coupling between individual Ca^{2+} channels. *Neuron* **9**, 197–207 (1992).
- McCleskey, E. W. et al. ω -Conotoxin: direct and persistent blockade of specific types of calcium channels in neurons but not muscle. *Proc. Natl Acad. Sci. USA* **84**, 4327–4331 (1987).
- McDonough, S. L., Swartz, K. J., Mintz, I. M., Polan, L. M. & Bean, B. P. Inhibition of calcium channels in rat central and peripheral neurons by ω -conotoxin MVIIIC. *J. Neurosci.* **16**, 2612–2623 (1995).
- Marrion, N. V. Calcineurin regulates M-channel model gating in sympathetic neurons. *Neuron* **16**, 163–173 (1996).
- Yue, D. T. & Marban, E. Permeation in the dihydropyridine-sensitive calcium channel. Multi-ion occupancy but no anomalous mole-fraction between Ba^{2+} and Ca^{2+} . *J. Gen. Physiol.* **95**, 911–939 (1990).
- Horn, R. & Lange, K. Estimating kinetic constants from single channel data. *Biophys. J.* **43**, 207–223 (1983).
- Yue, D., Backs, P. H. & Imredy, J. P. Calcium-sensitive inactivation in the gating of single calcium channels. *Science* **250**, 1735–1738 (1990).

Acknowledgements. We thank B. Hirschberg, J. Maylie, D. Shepherd and J. T. Williams for reading the manuscript, and D. Shepherd for cell preparation. This work was supported by the NIH (NINDS).

Correspondence and requests for materials should be addressed to N.V.M. (e-mail: N.V. Marrion@bris.ac.uk).

Mice without myoglobin

Daniel J. Garry*, George A. Ordway†, John N. Lorenz‡, Nina B. Radford*, Eva R. Chin*, Robert W. Grange†, Rhonda Bassel-Duby* & R. Sanders Williams*§

Departments of *Internal Medicine, †Physiology, and §Molecular Biology/Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

‡Department of Molecular and Cellular Physiology, University of Cincinnati, Cincinnati, Ohio 45267, USA

Myoglobin, an intracellular haemoprotein expressed in the heart and oxidative skeletal myofibres of vertebrates, binds molecular oxygen and may facilitate oxygen transport from erythrocytes to mitochondria, thereby maintaining cellular respiration during periods of high physiological demand^{1–10}. Here we show, however, that mice without myoglobin, generated by gene-knockout

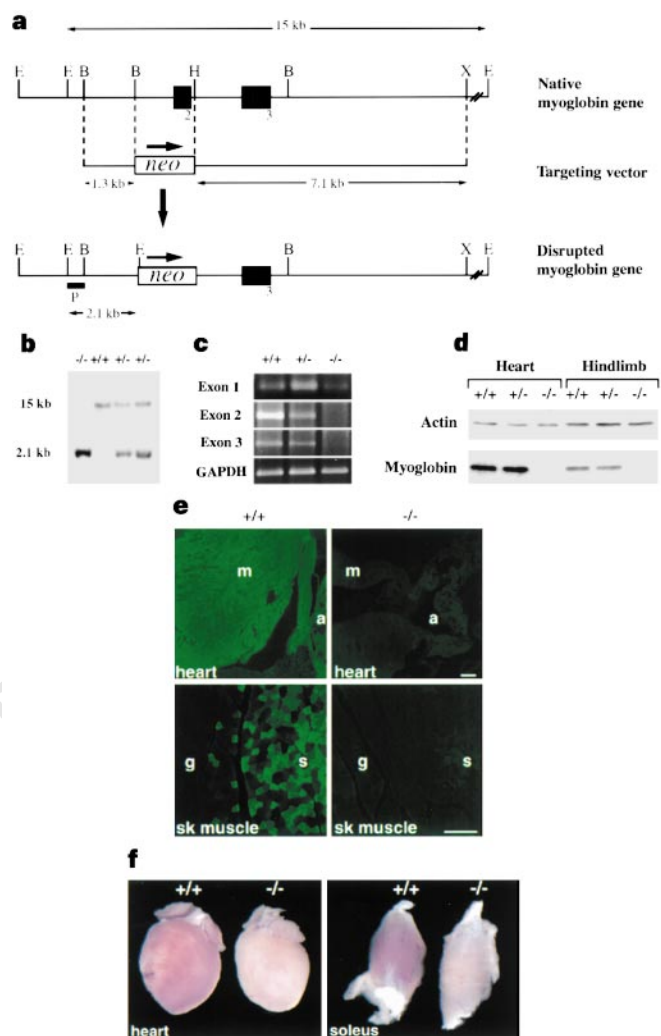


Figure 1 Targeted disruption of the myoglobin gene. **a**, The targeting vector was constructed by replacing a *Bam*HI–*Hin*dIII fragment containing exon 2 with a RNA polymerase II promoter driving the neomycin expression cassette¹⁴. Exons are shown as filled boxes and the position of the DNA segment used as a probe for Southern blot analysis is shown as P, *E*, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; X, *Xho*I. **b**, Southern blot analysis using *Eco*RI-digested genomic DNA. **c**, Reverse-transcription-PCR assay to detect transcripts produced from the native and targeted myoglobin alleles. Amplification products of the predicted sizes were produced with all three primer pairs using RNA derived from *myoglobin*^{+/+} and *myoglobin*^{−/−} animals. Transcripts derived from exon 1 continue to be produced in *myoglobin*^{−/−} mice, but longer transcripts that include sequences from exons 2 or 3 could not be detected. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) sequences were amplified in parallel to demonstrate integrity of the RNA in all samples. **d**, Immunoblot of protein extracts from hearts and hindlimb muscles. Myoglobin protein could not be detected in heart and hindlimbs of *myoglobin*^{−/−} mice. **e**, Immunolocalization of myoglobin in hearts or hindlimb skeletal (sk) muscles of *myoglobin*^{+/+} and *myoglobin*^{−/−} animals using a goat polyclonal antibody against myoglobin. Both atrial and ventricular myocytes of adult *myoglobin*^{+/+} mice express myoglobin, and expression in skeletal myofibres is limited to type I and IIa subtypes, as described³. In contrast, immunoreactive myoglobin could not be detected in sections of heart and skeletal muscles from *myoglobin*^{−/−} mice. The integrity of muscle proteins in tissue sections derived from *myoglobin*^{−/−} animals was confirmed by a positive immunostain for desmin (not shown). a, atrium; m, ventricular myocardium; s, soleus; g, gastrocnemius. Bar = 1,000 μ m (**a**, **b**) or 750 μ m (**c**, **d**). **f**, Gross appearance of perfused heart and soleus muscles, showing depigmentation in the absence of myoglobin.

technology, are fertile and exhibit normal exercise capacity and a normal ventilatory response to low oxygen levels (hypoxia). Heart and soleus muscles from these animals are depigmented, but function normally in standard assays of muscle performance *in vitro* across a range of work conditions and oxygen availability. These data show that myoglobin is not required to meet the metabolic requirements of pregnancy or exercise in a terrestrial mammal, and raise new questions about oxygen transport and metabolic regulation in working muscles.

Organisms as diverse as annelid worms, molluscs, flowering plants and protochordates express cytoplasmic haemoproteins (tissue haemoglobins) that bind oxygen reversibly¹, probably to facilitate oxygen transfer to sites of use². In vertebrates, only cardiac myocytes and a subset of mitochondria-rich skeletal myofibres (types I and IIa) express myoglobin^{3,4}, a monomeric haemoprotein that is distantly related to erythroid haemoglobins. Myoglobin concentrations in skeletal muscles are increased in mammals adapted for deep-sea diving⁵ or life at high altitude^{6,7}, or in response to sustained tonic muscle contractions^{8,9}.

Calculations of oxygen flux by simple diffusion suggest that, in the absence of myoglobin, oxygen pressure at the centre of myocytes working in normal oxygen (normoxic) levels would fall to levels insufficient to maintain mitochondrial oxidative phosphorylation¹. In support of this prediction, contractile function of cardiac or skeletal muscles is sensitive to chemical agents (nitrite, hydroxylamine or hydroperoxides) or carbon monoxide administered to reduce effective concentrations of oxymyoglobin¹⁰. The gene-knockout approach used here, however, avoids potential problems of specificity arising from the administration of chemical agents. The discovery of Antarctic icefishes lacking myoglobin¹¹ showed that vertebrates can survive in the absence of this specialized oxygen-binding protein, but no homeothermic organisms lacking myoglobin have been described so far.

Our gene-targeting strategy is shown in Fig. 1a, and a genotyping analysis of wild-type (*myoglobin*^{+/+}), heterozygous (*myoglobin*^{+/-}) and homozygous null (*myoglobin*^{-/-}) mice is shown in Fig. 1b. The targeting strategy deleted exon 2, which encodes amino acids 31–105 of the 154-residue myoglobin protein. The deleted region encodes the haem-binding domain, including the histidine residues at positions 65 and 94 that are essential to stabilize the haem group. Moreover, in the event of exon 1–3 splicing of primary transcripts derived from the targeted allele, the deletion of exon 2 would alter the reading frame and introduce a stop codon that would terminate translation after synthesis of only 39 amino acids. In the analysis of *myoglobin*^{-/-} mice, no RNA transcripts that include exon 2 or exon 3 sequences could be detected by a reverse-transcription/polymerase chain reaction (PCR) assay (Fig. 1c), or by *in situ* hybridization (results not shown). Likewise, neither full-length nor truncated forms of myoglobin protein could be detected in muscles of *myoglobin*^{-/-} mice, using two different polyclonal antibodies in immunoblot (Fig. 1d) and immunohistochemical (Fig. 1e) assays. Finally, we confirmed the null phenotype by the depigmentation observed in hearts or soleus muscles from animals homozygous for the targeted allele (Fig. 1f).

Mice lacking myoglobin showed no obvious phenotypic abnormalities during ambient activities. They gained weight at a rate indistinguishable from that of *myoglobin*^{+/-} or *myoglobin*^{+/+} littermates and reached sexual maturity at a comparable age. *Myoglobin*^{-/-} mice were fertile in matings with *myoglobin*^{+/-} or *myoglobin*^{+/+} animals. The histological appearance of sections of cardiac and skeletal muscles assessed by light microscopy with routine stains appeared normal. Among a colony that included more than 150 *myoglobin*^{-/-} animals up to 10 months of age, no animals showed overt signs of congestive heart failure, and the frequency of unexplained spontaneous deaths was comparable among all genotypes.

We subjected the animals to an exercise testing protocol¹² designed to assess endurance performance at an intensity of exercise

requiring high rates of oxygen flux. The exercise capacities of *myoglobin*^{+/+} and *myoglobin*^{-/-} mice were indistinguishable (Fig. 2a), and comparable to those of wild-type mice of this or other strains tested previously. Animals were also challenged by exposure to a hypoxic gas mixture containing 13.5% oxygen (100 torr; equivalent to an altitude of 3,700 m). *Myoglobin*^{-/-} mice survived this hypoxic challenge, and exhibited ventilatory responses during and following the period of reduced inspiratory oxygen tension that were comparable to those of *myoglobin*^{+/+} animals (Fig. 2b).

Studies of isolated cardiac and skeletal muscles using standard *in vitro* preparations support the conclusion that muscle performance remains at or near normal levels in the absence of myoglobin, across a range of work conditions and oxygen availability. The response of working heart preparations to increases in preload was normal in *myoglobin*^{-/-} hearts (Fig. 3a), even during graded reductions in oxygen tension sufficient to reduce stroke work by 40% in *myoglobin*^{+/+} hearts (Fig. 3b). Twitch properties of isolated soleus and extensor digitorum longus skeletal muscles were likewise normal, and *myoglobin*^{-/-} muscles fatigued at the same rate as *myoglobin*^{+/+} muscles in prolonged stimulation protocols under oxygenated or hypoxic conditions (Fig. 3c).

These data show that myoglobin is not essential for apparently normal cardiovascular and musculoskeletal function in a terrestrial, homeothermic mammal with high metabolic demands. More

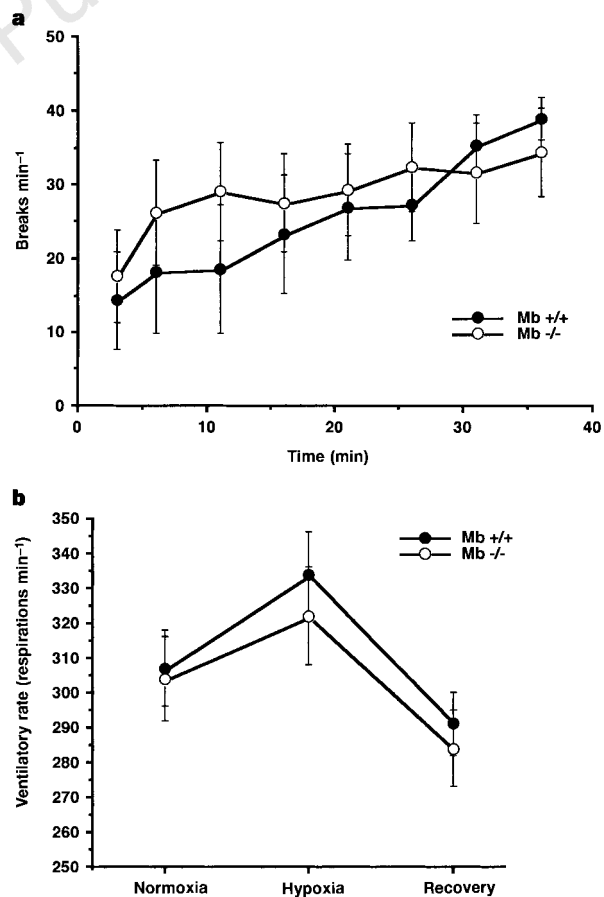


Figure 2 Physiological assessments of intact animals. **a**, Exercise performance of *myoglobin*^{+/+} and *myoglobin*^{-/-} animals. Data points depict mean values ($\pm$ s.e.m.) at the indicated times for five *myoglobin*^{+/+} and five *myoglobin*^{-/-} mice. Additional details of the exercise testing procedure have been published previously¹². **b**, Ventilatory responses to hypoxia in *myoglobin*^{+/+} and *myoglobin*^{-/-} animals. Data points depict mean values ($\pm$ s.e.m.) for ventilatory rate (respirations min⁻¹) before, during and after an 8-min exposure to the hypoxic environment (13.5% oxygen; equivalent to 100 torr or an altitude of 3,700 m) for seven *myoglobin*^{+/+} and six *myoglobin*^{-/-} mice. Respiratory rate was measured continuously as described¹⁷. Mb, myoglobin.

detailed studies of cardiac and skeletal muscle function and metabolism, or more extreme physiological challenges, may reveal abnormalities that were not detected by our initial tests, but our present, unexpected results raise new questions with respect to oxygen transport and muscle metabolism.

If myoglobin is not essential for routine activities of homeothermic animals, what evolutionary pressures have led to the conservation of the structure and expression patterns of myoglobin during vertebrate evolution? Perhaps exercise-testing protocols in the laboratory do not detect subtle differences in exercise performance that are sufficient to impart survival or reproductive advantages in the wild. On the other hand, evolutionary conservation of myoglobin structure and expression may be driven by regulatory functions other than bulk oxygen transport. For example, binding of nitric oxide to haemoglobin has vasoregulatory consequences¹³, indicating that myoglobin may also participate in nitric-oxide-dependent regulatory events that could be revealed by more detailed physiological investigations of myoglobin-deficient animals.

Is simple diffusion sufficient to facilitate high rates of oxygen transport to mitochondria of working myocytes in the absence of

myoglobin or do unknown redundant mechanisms exist? Animals that survive and thrive without myoglobin may do so by mounting adaptive responses that are unexpectedly robust. Preliminary studies indicate that muscles from *myoglobin*^{-/-} animals do not exhibit striking changes in mitochondrial content, vascular supply or myosin isoform expression, but other counter-regulatory phenomena, yet to be identified, may lead to preservation of muscle function in the absence of myoglobin. □

Methods

Disruption of the myoglobin gene. A myoglobin genomic clone was isolated by screening a P1 genomic library with a complementary DNA fragment which included exon 2 of the myoglobin gene. The targeting vector was constructed by replacing a *Bam*HI–*Hind*III fragment containing exon 2 with a RNA polymerase II promoter neomycin cassette¹⁴. The targeting vector consisted of a 1.3-kilobase (kb) *Bam*HI fragment of 5' myoglobin sequence, the neomycin cassette, a 7.1-kb fragment of 3' myoglobin sequence and a cassette encoding thymidine kinase drug resistance. Embryonic stem cell clones transfected with the targeting construct were obtained using a positive-negative selection strategy. Homologous recombinants were identified by digesting genomic DNA with *Eco*R1 and hybridizing the fragments with the 0.3-kb *Eco*R1/*Bam*HI 5' probe. Offspring were genotyped by Southern blot analysis.

RNA isolation and reverse-transcription-PCR. Total RNA was isolated from adult hearts of mice using the Tripure isolation kit (Boehringer). 10 µg total RNA was used in each reverse transcription reaction (Retro-Script; Ambion). Complementary DNA (5 µl) was then used as template for the PCR in a 50-µl reaction volume as described³. A single forward primer was positioned to include the initiation codon encoded by exon 1, and amplification by PCR (30 cycles) was done using reverse primers complementary to sequences from exons 1, 2 or 3.

Western blot analysis. Protein was extracted from heart and hindlimb muscles, separated by electrophoresis through a 12% SDS–polyacrylamide gel, transferred to nylon filters, and probed with a rabbit polyclonal anti-myoglobin serum (Dako) or a monoclonal anti-actin antibody (Amersham).

Histological analysis and immunohistochemistry. Cardiac or skeletal muscles were perfusion-fixed with 4% paraformaldehyde and embedded in paraffin; immunohistochemistry was done as described³. Primary antiserum included rabbit anti-myoglobin (Dako), goat anti-myoglobin (Organon Teknika), and monoclonal anti-desmin (Biogenix Lab). The secondary antiserum used in these studies was fluorescein isothiocyanate (FITC)-conjugated, species-specific IgG (Jackson ImmunoResearch). Immunostained sections were studied with a Bio-Rad MRC 1000 confocal microscope equipped with a krypton/argon laser (Bio-Rad Life Science Group).

Exercise treadmill study. Mice ran daily over 4 days on a motor-driven treadmill (Omni-Pacer LC4, Omnitech) equipped with shock bars with adjustable amperage and an infrared detection system, as described¹². The first 3 days constituted an acclimation period during which the speed and the total duration of exercise were increased incrementally. Formal testing was done on the fourth day. Exercise performance was assessed by the frequency at which animals blocked two infrared light barriers positioned at the rear of the exercise chamber (breaks min⁻¹) while running on a treadmill at a speed of 20 m min⁻¹ for the first 3 min, 25 m min⁻¹ for the next 3 min and 27 m min⁻¹ for the remaining 30 min at a 7% incline. Personnel supervising the exercise testing and the hypoxic challenge were blinded to the genotypes of the animals.

Isolated working heart perfusions. Working heart preparations were established and analysed as described¹⁵. Stroke work (mean aortic pressure multiplied by stroke volume, corrected for cardiac mass) at a constant perfusing pressure (80 cm H₂O) was analysed in response to changes in preload. Intrinsic heart rates were similar (<5% difference) in *myoglobin*^{+/+} and *myoglobin*^{-/-} hearts. Haemodynamic measurements were acquired after an initial stabilization period of 15 min, and after 15 min at each preload. The response to hypoxic conditions was evaluated by reducing the oxygen tension of the perfusate and pacing the hearts at a rate of 400 beats min⁻¹. Data were acquired at three intervals 5 min apart at each level of oxygen tension.

Skeletal muscle contractility measurements. Soleus and extensor digitorum muscles (*n* = 7 in each group) were isolated from anaesthetized mice and suspended in a jacketed organ bath containing an oxygenated Krebs

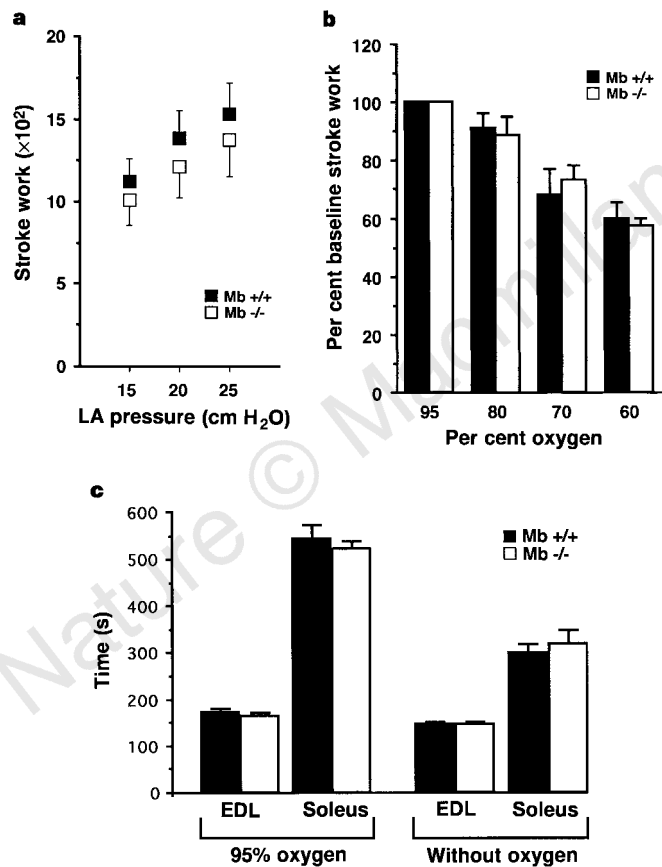


Figure 3 Physiological assessments of isolated muscle preparations. **a**, Working heart preparations were established and analysed as described¹⁵. Stroke work (mean aortic pressure multiplied by stroke volume, corrected for cardiac mass) at a constant perfusing pressure (80 cm H₂O) increased normally in response to changes in preload in *myoglobin*^{-/-} hearts. Intrinsic heart rates were similar (<5% difference) in both groups. Data points depict mean values (± s.e.m.) from four *myoglobin*^{+/+} and six *myoglobin*^{-/-} mice. LA, left atrial. **b**, With reductions in oxygen tension of the perfusate, stroke work declined similarly in *myoglobin*^{-/-} and *myoglobin*^{+/+} hearts. Data points depict mean values (± s.e.m.) from four *myoglobin*^{+/+} and five or six *myoglobin*^{-/-} mice. **c**, Soleus and extensor digitorum longus (EDL) muscles (*n* = 7 in each group) were isolated and suspended in a jacketed organ bath. Muscle length and stimulation voltage were adjusted to yield maximal isometric twitch force. The time to fatigue (70% decline in force) in response to intermittent tetani (350 ms at 100 Hz) was similar in *myoglobin*^{-/-} and *myoglobin*^{+/+} muscles under both well-oxygenated and hypoxic conditions.

solution at 25 °C and a Grass Ft.03 force transducer as described¹⁶. Muscles were stimulated using closely flanking platinum wire electrodes. We studied the time to fatigue (70% decline in force) in response to intermittent tetani (350 ms at 100 Hz) under normoxic and hypoxic conditions.

Received 25 June; accepted 8 September 1998.

1. Wittenberg, B. A. & Wittenberg, J. B. Transport of oxygen in muscle. *Annu. Rev. Physiol.* **51**, 857–878 (1989).
2. Wittenberg, J. B. & Wittenberg, B. A. Mechanisms of cytoplasmic hemoglobin and myoglobin function. *Annu. Rev. Biophys. Chem.* **19**, 217–241 (1990).
3. Garry, D. J. et al. Postnatal development and plasticity of specialized muscle fiber characteristics in the hindlimb. *Dev. Genet.* **19**, 146–156 (1996).
4. Parsons, W. J., Richardson, J. A., Graves, K. H., Williams, R. S. & Moreadith, R. W. Gradients of transgene expression directed by the human myoglobin promoter in the developing mouse heart. *Proc. Natl Acad. Sci. USA* **90**, 1726–1730 (1993).
5. Guyton, G. P. et al. Myoglobin saturation in free-diving Weddell seals. *J. Appl. Physiol.* **79**, 1148–1155 (1995).
6. Gimenez, M., Sanderson, R. J., Reiss, O. K. & Banchero, N. Effects of altitude on myoglobin and mitochondrial protein in canine skeletal muscle. *Respiration* **34**, 171–176 (1977).
7. Terrados, N. Altitude training and muscular metabolism. *Int. J. Sports Med.* **13** (Suppl.), 206–209 (1992).
8. Wittenberg, J. B. Myoglobin facilitated oxygen diffusion and the role of myoglobin in oxygen entry into muscle. *Physiol. Rev.* **50**, 559–636 (1970).
9. Underwood, L. E. & Williams, R. S. Pretranslational regulation of myoglobin gene expression. *Am. J. Physiol.* **252**, C450–C453 (1987).
10. Doeller, J. E. & Wittenberg, B. A. Myoglobin function and energy metabolism of isolated cardiac myocytes: effect of sodium nitrite. *Am. J. Physiol.* **261**, H53–H62 (1991).
11. Sidell, B. D. et al. Variable expression of myoglobin among the hemoglobinless Antarctic icefishes. *Proc. Natl Acad. Sci. USA* **94**, 3420–3424 (1997).
12. Fewell, J. G. et al. A treadmill exercise regimen for identifying cardiovascular phenotypes in transgenic mice. *Am. J. Physiol.* **273**, H1595–H1605 (1997).
13. Gow, A. J. & Stamler, J. S. Reactions between nitric oxide and haemoglobin under physiological conditions. *Nature* **391**, 169–173 (1998).
14. Ishibashi, S. et al. Hypercholesterolemia in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery. *J. Clin. Invest.* **92**, 883–893 (1993).
15. Grupp, I. L., Subramaniam, A., Hewett, T., Robbins, J. & Grupp, G. Comparison of normal, hypodynamic and hyperdynamic mouse hearts using isolated work-performing heart preparations. *Am. J. Physiol.* **265**, H1401–H1410 (1993).
16. Ho, C. S., Grange, R. W. & Joho, R. H. Pleiotropic effects of a disrupted K⁺ channel gene: reduced body weight, impaired motor skill and muscle contraction, but no seizures. *Proc. Natl Acad. Sci. USA* **94**, 1533–1538 (1997).
17. Hamelmann, E. et al. Noninvasive measurement of airway responsiveness in allergic mice using barometric plethysmography. *Am. J. Respir. Crit. Care Med.* **156**, 766–775 (1997).

Acknowledgements. We thank J. Shelton, M. Fina, C. Storey and A. Richman for technical assistance. This work was supported by grants from the NIH (to R.S.W., R.B.-D. & D.J.G.) and American Heart Association (to N.B.R.).

Correspondence and requests for materials should be addressed to R.S.W. (e-mail: williams@ryburn.swmed.edu).

Role of Ca²⁺/K⁺ ion exchange in intracellular storage and release of Ca²⁺

Thien Nguyen[†], Wei-Chun Chin^{*†} & Pedro Verdugo^{†‡}

Departments of [†]Bioengineering and [‡]Internal Medicine, University of Washington, Box 357962, Seattle, Washington 98195, USA

* These authors contributed equally to this work.

Although fluctuations in cytosolic Ca²⁺ concentration have a crucial role in relaying intracellular messages in the cell¹, the dynamics of Ca²⁺ storage in and release from intracellular sequestering compartments remains poorly understood. The rapid release of stored Ca²⁺ requires large concentration gradients that had been thought to result from low-affinity buffering of Ca²⁺ by the polyanionic matrices within Ca²⁺-sequestering organelles². However, our results here show that resting luminal free Ca²⁺ concentration inside the endoplasmic reticulum and in the mucin granules remains at low levels (20–35 µM). But after stimulation, the free luminal [Ca²⁺] increases, undergoing large oscillations, leading to corresponding oscillations of Ca²⁺ release to the cytosol. These remarkable dynamics of luminal [Ca²⁺] result from a fast and highly cooperative Ca²⁺/K⁺ ion-exchange process rather than from Ca²⁺ transport into the lumen. This common paradigm for Ca²⁺ storage and release, found in two different Ca²⁺-sequestering

organelles, requires the functional interaction of three molecular components: a polyanionic matrix that functions as a Ca²⁺/K⁺ ion exchanger, and two Ca²⁺-sensitive channels, one to import K⁺ into the Ca²⁺-sequestering compartments, the other to release Ca²⁺ to the cytosol.

Ion exchange has a crucial role in the release of histamine from isolated mast-cell granules³. The polymer gel phase transition that results in the quick exocytotic decondensation of the mucin secretory matrix is also governed by intragranular-Ca²⁺/extracellular-Na⁺ ion exchange⁴. Our present observations further broaden the role of ion exchange into the dynamics of Ca²⁺ storage in and release from Ca²⁺-signalling intracellular compartments.

We performed experiments with fluorescent Ca²⁺ probes and thin optical sectioning techniques in both intact cells and exposed organelles⁵. To broaden the dynamic range of detection of [Ca²⁺], we loaded goblet cells with two acetyl methyl esters of the Ca²⁺-sensitive dyes rhod-2 ($K_d = 570$ nM) and calcium orange-5N ($K_d = 20$ µM) (see Methods). The endoplasmic reticulum of ciliated cells was loaded with either calcium orange-5N, for single-wavelength measurements of [Ca²⁺] ([Ca²⁺]_{ER}), or with Mag-Indo-1 ($K_d = 35$ µM), for ratiometric detection of [Ca²⁺]_{ER}. To minimize photobleaching of the Ca²⁺ probes while focusing occurred, the granules and the endoplasmic reticulum were double-labelled with quinacrine^{5,6} and 3,3'-dihexyloxycarbocyanine iodide (DiOC₆(3)) (ref. 7) respectively (see Methods). The [Ca²⁺] was calculated from measurements of fluorescence intensity performed on 200-nm optical sections of fluorescent images⁵.

Exposure of intact goblet cells to 100 µM ATP triggered a burst of Ca²⁺-induced fluorescence oscillations inside and outside the granules with a period of 10–15 seconds. The descending phase of each intraluminal oscillation coincided with a corresponding increase in fluorescence in the neighbouring region outside the granule ($n = 10$) (Fig. 1a) and the subsequent formation of a Ca²⁺-induced fluorescence wave that spread ~2 µm into the cytosol (Fig. 2b). As this localized extragranular wave dissipated, the intragranular fluorescence increased again in the next cycle. The reciprocal changes between intragranular and cytosolic fluorescence suggest that the oscillations of cytosolic fluorescence are derived from the release of luminal Ca²⁺.

In isolated mucin granules incubated in intracellular buffer (see Methods), single-wavelength fluorescence measurements indicated a stable resting intragranular [Ca²⁺] ([Ca²⁺]_G) of 24 ± 0.8 µM ($n = 10$). The addition of 3 µM inositol-1,4,5-trisphosphate (InsP₃) resulted in [Ca²⁺]_G oscillations similar to those produced by ATP in intact cells (Fig. 1b). Unlike observations in intact cells, the binding of isolated granules to the coverslip confines the granules to a single focal plane, avoiding motion artefacts and allowing the monitoring of [Ca²⁺]_G in several granules simultaneously. Observations of such sets of granules revealed that not all the isolated mucin granules respond to InsP₃ stimulation. The release of Ca²⁺ from isolated granules was verified by equilibrating them in intracellular buffer containing non-permeant dextran-conjugated calcium crimson⁸. Again, the descending phase of each oscillation of [Ca²⁺]_G coincided with a corresponding increase in [Ca²⁺] in the neighbouring region outside the granule. The reciprocal changes in [Ca²⁺] inside and outside the granules provide objective evidence that InsP₃ induces oscillations in [Ca²⁺]_G and a corresponding periodic Ca²⁺ release from the granules (Fig. 1b).

Thin optical sections of the endoplasmic reticulum reveal a network of cisternal segments (Fig. 2e). Exposure of intact ciliated cells to 100 µM ATP triggered a sequence of [Ca²⁺]-induced oscillations inside the endoplasmic reticulum and a corresponding phase-shifted train of fluorescence oscillations in the neighbouring cytosol (Fig. 3a). Measurements of [Ca²⁺]_{ER} in permeabilized epithelial ciliated cells loaded with the Ca²⁺-sensitive dye calcium orange-5N yielded a stable resting [Ca²⁺]_{ER} of 28 ± 5 µM ($n = 36$ cells from 36 cultures). Verification with ratiometric measurements

with Mag-Indo-1 yielded a higher but not significantly different $[Ca^{2+}]_{ER}$ of $34 \pm 8 \mu M$ ($n = 80$ cells from 10 tissue cultures) (see Methods). Direct stimulation by $InsP_3$ ($3 \mu M$) of the endoplasmic reticulum of permeabilized ciliated cells equilibrated in ATP-free intracellular buffer caused a train of $[Ca^{2+}]_{ER}$ oscillations (Fig. 3b). Here, quenching of cytosolic fluorescence by Mn^{2+} prevented the observation of Ca^{2+} release from permeabilized ciliated cells (see Methods).

The $InsP_3$ -induced fluctuations of $[Ca^{2+}]_{ER}$ and $[Ca^{2+}]_G$ remained unaffected when permeabilized ciliated cells or mucin granules were stimulated with $3 \mu M$ $InsP_3$ in the presence of 100 nM thapsigargin, a blocker of the sarcoplasmic/endoplasmic-reticulum Ca^{2+} -ATPase (SERCA)¹⁰, or in the presence of $2 \mu M$ ruthenium red, a potent

blocker of the mitochondrial Ca^{2+} uniporter¹¹ (data not shown). These results suggest that the $[Ca^{2+}]_{ER}$ and $[Ca^{2+}]_G$ fluctuations did not result from uptake by SERCA or mitochondrial activity. Conversely, the endoplasmic reticulum's and the granule's response to $InsP_3$ was abolished when the preparations were pretreated with an intracellular buffer containing heparin ($100 \mu g ml^{-1}$), a blocker of the $InsP_3$ receptor¹², or apamin (100 nM), a specific blocker of small-conductance Ca^{2+} -activated K^+ (SK_{Ca}) channels¹³ (data not shown). The blocking effect of heparin confirms the presence of an $InsP_3$ receptor in the membrane of these organelles and the blockage produced by apamin strongly suggests the existence of an apamin-sensitive Ca^{2+} -activated K^+ channel (ASK_{Ca}) in the endoplasmic reticulum and the mucin granules.

Additional observations further support the presence of a Ca^{2+} -sensitive K^+ channel in the endoplasmic reticulum and the granule. For instance, adding tetraethylammonium (TEA) (20 mM), a general K^+ -channel blocker¹³, or replacing K^+ in the intracellular buffer by the monovalent organic cation *N*-methyl-D-glucamine (NMG) at constant ionic strength, suppresses the effect of $InsP_3$ in the endoplasmic reticulum and granules (data not shown). Moreover, the $InsP_3$ -induced $[Ca^{2+}]$ oscillations were diminished or abolished by lowering $[Ca^{2+}]$ to below 100 nM in the intracellular buffer (data not shown). These results are consistent with previous demonstrations that the Ca^{2+} content of secretory granules decreases after stimulation whereas the K^+ content increases¹⁴, and the $InsP_3$ -induced Ca^{2+} release from endoplasmic reticulum vesicles requires the inflow of K^+ into the cisternae via an SK_{Ca} channel¹⁵; however, the evidence presented here is the first report of an ASK_{Ca} channel in secretory granules.

The inflow of K^+ into the endoplasmic reticulum has been interpreted as a requirement to maintain electroneutrality when Ca^{2+}

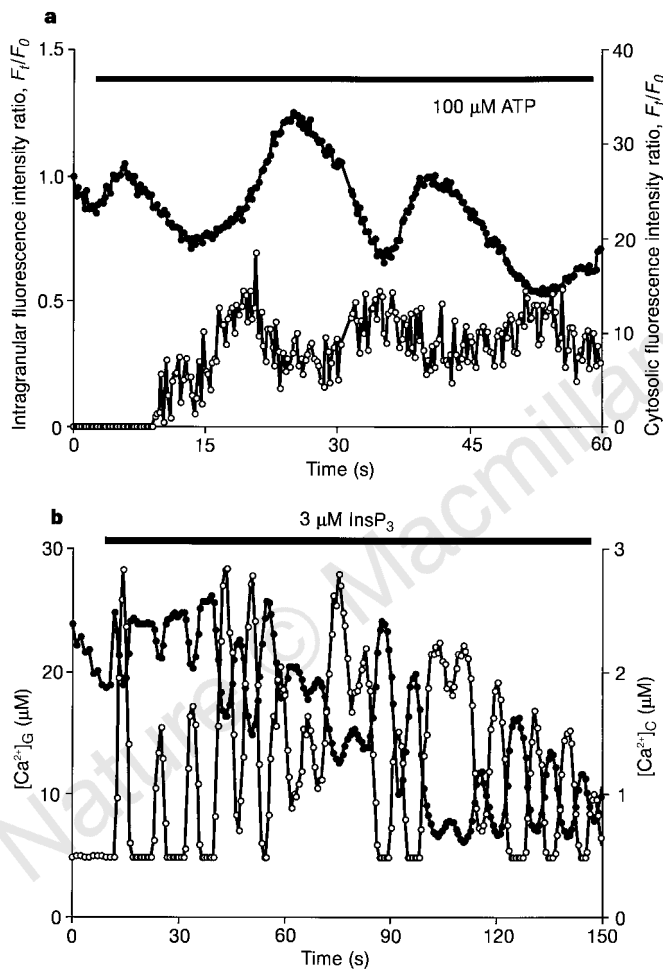


Figure 1 Oscillations of $[Ca^{2+}]$ in intact goblet cells and in isolated secretory vesicles. **a**, Measurements of changes of $[Ca^{2+}]$ -induced fluorescence as a function of time both inside (filled circles) and outside (open circles) a mucin granule of an intact goblet cell exposed to $100 \mu M$ ATP ($n = 10$). The intraluminal fluorescence oscillates with a quasi-sinusoidal pattern. A corresponding train of cytosolic oscillations of fluorescence, out of phase with the intraluminal oscillations, occurs in the cytosol immediately adjacent to the granule. **b**, $[Ca^{2+}]$ oscillations inside (filled circles) and outside (open circles) an isolated mucin granule. The granule was loaded with calcium orange-5N to detect $[Ca^{2+}]_G$, then suspended in an intracellular buffer containing $3 \mu M$ $InsP_3$ and $10 \mu g ml^{-1}$ dextran-conjugated calcium crimson to detect Ca^{2+} release ($n = 10$). As in the intact cell, the release of Ca^{2+} from the granules generates a corresponding wave of fluctuating $[Ca^{2+}]$ outside the granule which is about 180° out of phase with the $[Ca^{2+}]_G$ oscillation. The difference in frequency of $[Ca^{2+}]$ oscillations in intact cells (0.06 Hz, Fig. 1a) vs isolated granules (0.13 Hz) might result from differences in Ca^{2+} buffering and/or Ca^{2+} diffusion rates between the cytosol and the dextran-containing intracellular buffer (see Methods).

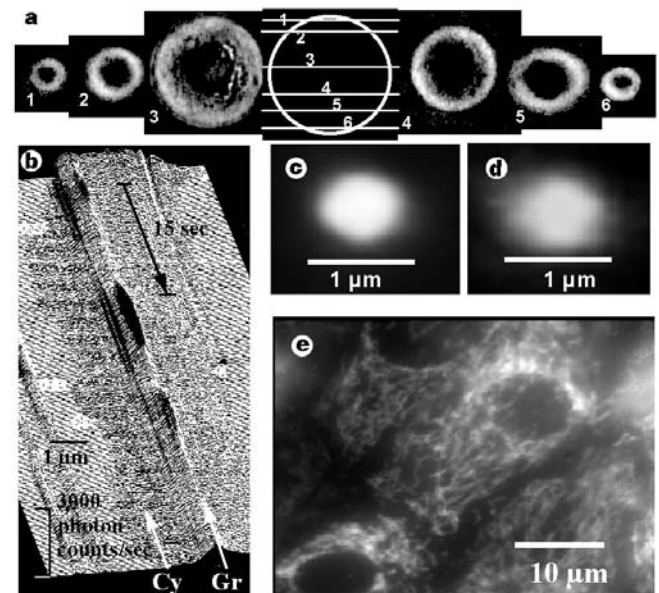


Figure 2 Validation of the Ca^{2+} detection and optical sectioning methods. **a**, the validation of the optical deconvolution software was conducted by performing optical sections of a $2\text{-}\mu m$ fluorescently coated transparent microsphere (Polysciences). The deconvoluted images of six horizontal sections provide the expected toroidal images yielding a z-axis resolution of $\sim 160\text{--}200$ nm. **b**, Three-dimensional plot of fluorescence intensity as a function of time in a 60-s line-stack of a goblet cell after stimulation with $100 \mu M$ ATP. This back-to-front view shows that $[Ca^{2+}]$ oscillations take place both inside and outside the granule, and that the oscillations of $[Ca^{2+}]_G$ and $[Ca^{2+}]_C$ are out of phase. Notice the formation of a tidal wave of Ca^{2+} that spreads and rebounds about $2 \mu m$ into or from the cytosol. **c**, Optical sections of two mucin secretory granules labelled with quinacrine. **d**, Optical sections of two mucin secretory granules labelled with serotonin. **e**, A 200-nm optical section of a ciliated cell illustrating the $DiOC_6(3)$ -labelled endoplasmic reticulum cisternae.

is released from the cisternae¹⁶. Here we show that electroneutrality notwithstanding, the inflow of K^+ into the endoplasmic reticulum or the secretory granules results in a marked increase in luminal $[Ca^{2+}]$ ($[Ca^{2+}]_L$) that cannot be explained by electroneutrality.

A family of Ca^{2+} -binding glycoproteins inside the endoplasmic reticulum, including calsequestrin and calreticulin, is believed to keep the free $[Ca^{2+}]_{ER}$ buffered at a steady high concentration (100 μM to 300 mM), providing about three orders of magnitude of $[Ca^{2+}]$ gradient across the endoplasmic reticulum membrane during Ca^{2+} release^{2,17,18}. However, as shown here, the resting $[Ca^{2+}]_{ER}$ remains at only 25–35 μM , but increases 10-fold after stimulation with $InsP_3$. This marked $InsP_3$ -induced increase in $[Ca^{2+}]_{ER}$ does not require uptake by SERCA. An alternative explanation for these unexpected results is that the glycoprotein matrix in the endoplasmic reticulum functions as an ion-exchange resin^{2,18} rather than as a simple buffer. Several lines of evidence support this notion. For instance, purified calsequestrin and calreticulin—the main Ca^{2+} -buffers in the endoplasmic reticulum and sarcoplasmic reticulum—show a 6- to 10-fold decrease in Ca^{2+} binding when the K^+ concentration in the incubating medium increases from 20 to

150 mM at constant ionic strength¹⁹. This suggests that the matrix of the endoplasmic reticulum can operate as a Ca^{2+}/K^+ ion-exchanger in which the Ca^{2+} binding is modulated by inflow of K^+ into the endoplasmic reticulum.

To study directly the Ca^{2+}/K^+ ion exchange inside the endoplasmic reticulum, permeabilized ciliated cells were loaded with calcium orange-5N and equilibrated in ATP-free intracellular buffer containing 100 nM Ca^{2+} , 100 $\mu g\ ml^{-1}$ heparin and 100 nM apamin^{12,13}. Under these conditions the resting $[Ca^{2+}]_{ER}$ remains stable, suggesting that the $InsP_3$ receptor channel, the ASK_{Ca} channel and the SERCA pump of the endoplasmic reticulum were rendered inoperative. Nigericin (10 μM), a K^+-H^+ ionophore, was added to the intracellular buffer to equilibrate K^+ and H^+ across the endoplasmic reticulum membrane¹⁵ (see Methods). The $[K^+]$ in the buffer was varied from 1 to 150 mM at pH 5.5. The ionic strength and osmolality of the medium were kept constant by adjusting the concentration of NMG in the intracellular buffer. Increasing $[K^+]$

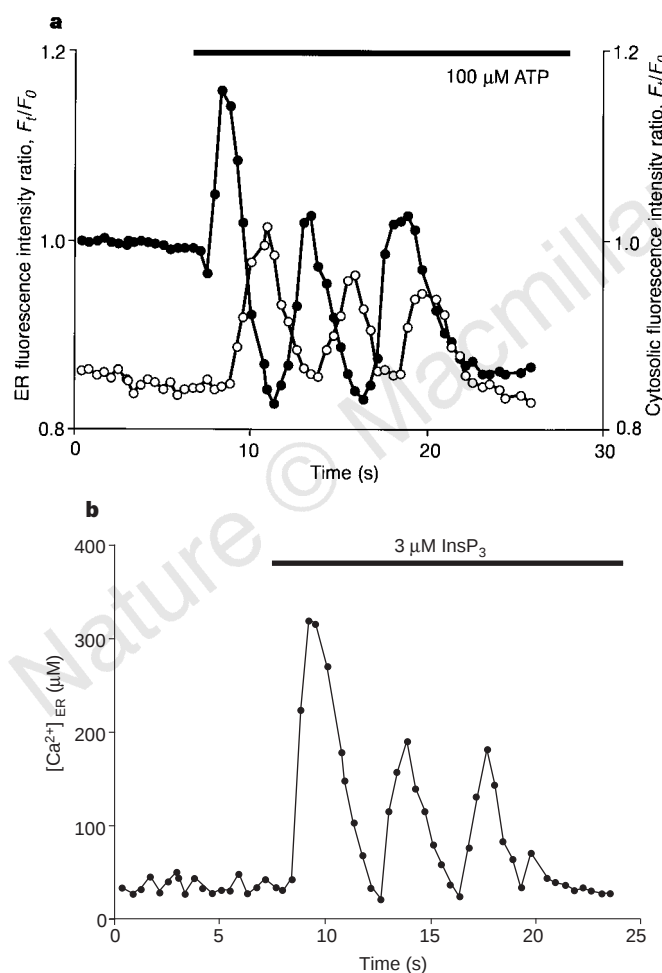


Figure 3 Oscillations of $[Ca^{2+}]_C$ and $[Ca^{2+}]_{ER}$ in intact and permeabilized ciliated cells. **a**, Measurements of fluctuations in $[Ca^{2+}]$ -induced fluorescence as a function of time inside (filled circles) and outside (open circles) the endoplasmic reticulum (ER) of ciliated cells exposed to 100 μM ATP ($n = 10$). Note that the intraluminal oscillations in fluorescence have a periodic pattern with a frequency of ~ 0.2 Hz. A similar sequence of cytosolic oscillations in $[Ca^{2+}]$ -induced fluorescence, out of phase with the intraluminal oscillations, propagates locally in the cytosol adjacent to the endoplasmic reticulum. **b**, Effect of 3 μM $InsP_3$ on $[Ca^{2+}]_{ER}$ in a permeabilized ciliated cell equilibrated in intracellular buffer ($n = 8$). Note that the frequency of $[Ca^{2+}]_{ER}$ oscillations in this case is slightly higher (0.3 Hz) than in intact cells, and that Mn^{2+} quenching prevented the detection of $[Ca^{2+}]_C$.

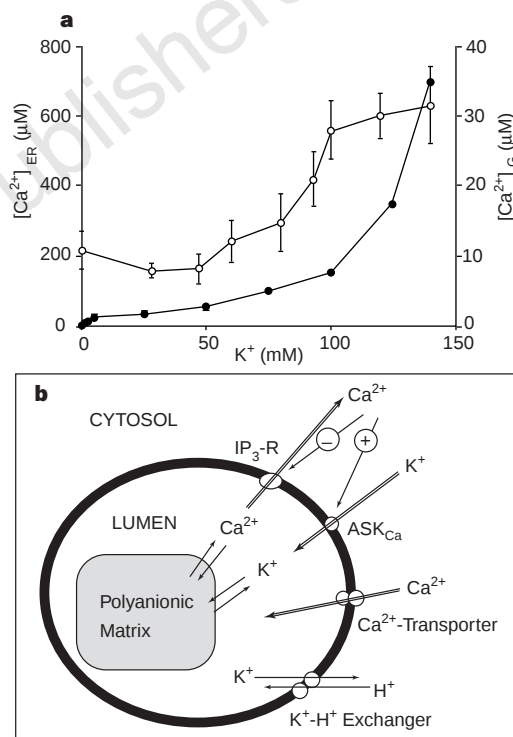


Figure 4 Ca^{2+}/K^+ ion-exchange in the ER and secretory vesicles. **a**, Effect of increasing K^+ concentration on $[Ca^{2+}]_{ER}$ (filled circles) and $[Ca^{2+}]_G$ (open circles) in permeabilized ciliated cells and isolated mucin granules equilibrated in ATP-free intracellular buffer (100 $\mu g\ ml^{-1}$ heparin, 100 nM apamin and 10 μM nigericin at pH 5.5 (10 mM MES)). The inflow of K^+ to the lumen results in an increase in $[Ca^{2+}]_{ER}$ or $[Ca^{2+}]_G$. Each point corresponds to the mean $\pm$ s.d. of 10 experiments. **b**, Dynamics of intracellular Ca^{2+} storage and release. This model assumes two pools of Ca^{2+} inside the lumen of the mucin granule or the ER: a pool of Ca^{2+} bound to the matrix and a free Ca^{2+} pool¹⁷. It further assumes that the $InsP_3$ receptor (IP_3 -R) (Ca^{2+} channel) and the ASK_{Ca} are in close proximity on the surface of these organelles. When $InsP_3$ triggers the opening of the $InsP_3$ receptor, the initial outflow of Ca^{2+} to the cytosol is driven by the resting $[Ca^{2+}]_{ER}$ or $[Ca^{2+}]_G$. Calcium release causes an increased $[Ca^{2+}]_C$ in the vicinity of the $InsP_3$ -receptor channel, triggering the opening of the ASK_{Ca} and the inflow of K^+ into the lumen^{13,14}. Inflow of K^+ serves two purposes: one is to balance electroneutrality and the other is to mobilize matrix-bound Ca^{2+} via ion exchange, producing an increase in $[Ca^{2+}]_{ER}$ or $[Ca^{2+}]_G$. The increase in intraluminal $[Ca^{2+}]$ promotes Ca^{2+} outflow. The Ca^{2+} release ultimately decreases the concentration of luminal Ca^{2+} and increases local $[Ca^{2+}]_C$. The increase in $[Ca^{2+}]_C$ results in closing of the $InsP_3$ receptor (Ca^{2+} channel)²⁸, the opening of the ASK_{Ca} channel and the initiation of a new cycle. Reloading of luminal Ca^{2+} would result from the SERCA pump in the endoplasmic reticulum, and from the H^+/Ca^{2+} exchanger or a Ca^{2+} -ATPase for secretory granules²⁷.

caused a nonlinear elevation of one to two orders of magnitude in $[Ca^{2+}]_{ER}$ with a steep sigmoidal increase in $[Ca^{2+}]_{ER}$ at 70–150 mM $[K^+]$ (Fig. 4a). These results agree with observations that nigericin can reverse the blockade of $InsP_3$ -induced Ca^{2+} release produced by apamin in endoplasmic reticulum vesicles¹⁵. However, our results further suggest that inflow of K^+ into the endoplasmic reticulum not only serves to maintain electroneutrality, but also causes unbinding of Ca^{2+} from the endoplasmic reticulum matrix.

Mucins, the main polymeric component in the mucous granule matrix, also have ion-exchange properties. Monovalent cations can exchange and release bound Ca^{2+} from mucin polymers in solution²⁰. To investigate Ca^{2+}/K^+ ion exchange in mucin granules, isolated granules loaded with calcium orange-5N were equilibrated in intracellular buffer containing 10 μ M nigericin¹⁵ while the $[K^+]$ was increased, as described earlier (see Methods). As in the endoplasmic reticulum, increasing the $[K^+]$ caused a nonlinear elevation in $[Ca^{2+}]_G$. Thus, the influx of K^+ into mucin granules could serve to readily mobilize Ca^{2+} bound to the mucin matrix, thereby increasing $[Ca^{2+}]_G$ (Fig. 4a). As in the endoplasmic reticulum, intragranular Ca^{2+}/K^+ ion exchange would serve to amplify the granule–cytosol Ca^{2+} diffusion gradient, also maintaining electroneutrality across the granular membrane. The mechanism of $[Ca^{2+}]$ oscillation in Ca^{2+} -storage compartments has been thought to require both a channel to release Ca^{2+} ($InsP_3$ receptor) and a Ca^{2+} -uptake system to import Ca^{2+} into the Ca^{2+} -sequestering compartments²¹. The lack of evidence for an active Ca^{2+} -uptake system in secretory granules has been thought to rule out the idea that $[Ca^{2+}]$ could increase inside secretory granules²¹. However, as shown here, free $[Ca^{2+}]_G$ can increase via Ca^{2+}/K^+ ion exchange without Ca^{2+} uptake.

Local fluctuations in cytosolic $[Ca^{2+}]$ ($[Ca^{2+}]_C$) are believed to be crucial for exocytosis: in the docking of the secretory granules, the fusion of plasma–granule membranes and the formation of the secretory pores²². Secretory granules contain large amounts of bound Ca^{2+} (ref. 14). Mobilization of a fraction of this bound Ca^{2+} pool could account for the local increase in $[Ca^{2+}]_C$ observed during exocytosis^{22,23}. The evidence presented here shows that the mucin granules of goblet cells are probably the source of the highly localized $[Ca^{2+}]_C$ increases that precede degranulation. Our results agree with previous observations that isolated pancreatic zymogen granules release Ca^{2+} when exposed to $InsP_3$ (ref. 23). These observations have been disputed, assigning the release of Ca^{2+} to contamination with endoplasmic reticulum, mitochondria or nuclei in the isolated granule preparation²⁴. Here, thin optical sectioning enabled reliable measurements of fluorescence inside and outside the mucin granule allowing unequivocal interpretation of our results (Fig. 2a–d). Three additional independent lines of evidence confirm that these vesicles are indeed mucin granules and that the observed $[Ca^{2+}]$ oscillations arise from the lumen of these granules and not from contamination with endoplasmic reticulum, mitochondria or nuclei. First, mucin secretory vesicles, like other secretory granules, were found to take up serotonin and quinacrine readily^{6,25} (Fig. 2c, d). Second, disruption of the granule membrane by detergent treatment resulted in rapid swelling of the mucin matrix which followed the characteristic first-order polymer-gel swelling kinetics previously observed during exocytosis of mucin secretory granules⁴ (data not shown). Last, the insensitivity of the $InsP_3$ -induced $[Ca^{2+}]$ oscillations to ruthenium red (2 μ M) indicates that the $[Ca^{2+}]$ oscillations that we observed arise from the lumen of the granule rather than from contamination with endoplasmic reticulum or mitochondria¹¹.

The presence of an $InsP_3$ receptor (Ca^{2+} channel) and a ASK_{Ca} (K^+ channel) with opposite conductance sensitivities to Ca^{2+} , allow the endoplasmic reticulum as well as the secretory granule to function as a Ca^{2+} oscillator regulated by $InsP_3$. The endoplasmic reticulum (or mucin granule) can generate periodic fluctuations in intraluminal $[Ca^{2+}]$ that result in the oscillatory release of Ca^{2+} into the

cytosol, and local $[Ca^{2+}]_C$ oscillations that precede the ATP-induced increase of ciliary beating²⁶ or the purinergic-induced degranulation in goblet cells²⁷. These results agree with the idea that the oscillatory nature of $[Ca^{2+}]_C$ rather than the absolute level of $[Ca^{2+}]_C$ might transduce intracellular signals¹, making the granule an excellent oscillator for periodic Ca^{2+} release.

A working model of the endoplasmic reticulum (granule)–cytosol Ca^{2+} exchange dynamics is illustrated in Fig. 4b. The binding of $InsP_3$ to its receptor results in the release of Ca^{2+} to the cytosol. This initial Ca^{2+} release is driven by the resting levels of $[Ca^{2+}]_L$ ($[Ca^{2+}]_{ER}$ or $[Ca^{2+}]_G$). The local release of Ca^{2+} results in $[Ca^{2+}]_C$ elevation in the neighbourhood of the $InsP_3$ receptor, opening the nearby ASK_{Ca} channels¹³. The influx of K^+ into the lumen raises the free $[Ca^{2+}]_L$ via a Ca^{2+}/K^+ ion exchange in the luminal matrix and markedly amplifies the lumen/cytosol diffusional gradient for local Ca^{2+} release. An increased release of Ca^{2+} from the endoplasmic reticulum or the granule results in decreasing $[Ca^{2+}]_L$ and generates a corresponding localized domain of high $[Ca^{2+}]_C$ that leads to inactivation of the $InsP_3$ receptor (ref. 28) while the ASK_{Ca} remains open, leading to increasing $[Ca^{2+}]_L$. As the local increase in $[Ca^{2+}]_C$ dissipates by diffusion or by cytosolic Ca^{2+} buffering, the ASK_{Ca} inactivates and the $InsP_3$ receptor reactivates, initiating a new cycle of Ca^{2+} release; thus, the cycle repeats. The present hypothesis offers several advantages for internal Ca^{2+} stores. Under resting conditions, $[Ca^{2+}]_L$ remains low (20–35 μ M), providing an important thermodynamic advantage for Ca^{2+} uptake, yet still maintaining a substantial diffusional drive to start Ca^{2+} release. The K^+ inflow modulates $[Ca^{2+}]_{ER}$ via Ca^{2+}/K^+ ion exchange, and counterbalances the build-up of a negative membrane potential that would otherwise oppose Ca^{2+} release¹⁶. It also modulates $[Ca^{2+}]_{ER}$ via Ca^{2+}/K^+ ion exchange. Thus, a large reservoir of bound Ca^{2+} can be quickly accessed for Ca^{2+} signalling without the thermodynamic disadvantage of maintaining a high $[Ca^{2+}]$ gradient and the potential for precipitation of intraluminal Ca^{2+} . □

Methods

Goblet cell culture and dye loading. Goblet cells from 25 tissue cultures grown from the trachea of five New Zealand white rabbits were used in these studies. The primary tissue-culturing protocol and the immunocytochemical identification of cultured respiratory goblet cells have been reported previously²⁹. Intact goblet cells were equilibrated at 37°C for 30 min in Hanks' solution containing 0.1% dimethylsulphoxide (DMSO), 5 μ M rhod-2 acetoxymethyl ester (a high-affinity Ca^{2+} -probe, $K_d = 570$ nM) and 5 μ M calcium orange-5N acetoxymethyl ester (a low-affinity Ca^{2+} -probe, $K_d = 20$ μ M) (Molecular Probes, Eugene, Oregon). This combination of two fluorescent probes yields measurements of a broad range of $[Ca^{2+}]_G$ and $[Ca^{2+}]_C$. Loading of the probes inside the secretory granules was verified by co-localization by using double-labelling of the granules with quinacrine (10 μ M) (Fig. 2c)^{5,6,25}. After dye loading, the cells were washed and equilibrated in Hanks' solution for 30 min. Because the excitation wavelengths of quinacrine and the Ca^{2+} probes are different (488 and 525 nm respectively), the double-labelling strategy allows the selection of image planes containing several granules without photobleaching the Ca^{2+} probes. After a selected cluster of granules inside a cell was identified and focused with the quinacrine label, the filter cube was changed to excite (525 nm) and collect the emission of rhod-2 and calcium orange-5N (580 nm) while the cells were stimulated with 100 μ M ATP²⁸.

Isolation and dye loading of mucin granules. Granules were isolated by following a procedure published elsewhere³. Briefly, the cultured goblet cells were double-labelled with quinacrine and calcium orange-5N acetoxymethyl ester. While the cells were suspended in a buffer that mimicked the intracellular environment (140 mM potassium glutamate, 500 nM Ca^{2+} (EGTA-buffered), 20 mM Tris pH 7.6), the cell membrane and cytoskeleton were disrupted by sonication. The granules were then separated by centrifugation and resuspended in intracellular buffer containing 10 μ g ml⁻¹ dextran-conjugated calcium crimson, a non-permeant Ca^{2+} probe⁸ ($K_d = 185$ nM; Molecular Probes). Aliquots of the granule suspension were then allowed to attach to poly-lysine-coated glass

chambers and mounted on the stage of the microscope. Oscillations of $[Ca^{2+}]_G$ and Ca^{2+} release in or from isolated granules were elicited by exposure to 3 μM $InsP_3$ (ref. 27). In isolated granules, heparin (100 $\mu g\ ml^{-1}$), apamin (100 nM), TEA (20 M), thapsigargin (100 nM) and ruthenium red (2 μM) were used to block the $InsP_3$ receptor, the SK_{Ca} channels, the SERCA and the mitochondrial Ca^{2+} -uniporter^{10–13}. NMG was used to replace K^+ in the K^+ -free intracellular buffer.

Ciliated cell culture and permeabilization protocols. Confluent monolayers of mucus-free ciliated cells from the trachea of five New Zealand white rabbits were grown in primary cultures, by following a protocol described previously²⁶. Once loaded with fluorescent Ca^{2+} probes (calcium orange-5N or Mag-Indo-1), the ciliated cells were permeabilized by a brief (4–8-min) treatment with 0.0025% digitonin. As soon as ciliary beating had stopped, the permeabilized cells were washed and equilibrated in an intracellular buffer containing 140 mM potassium glutamate, 3 mM ATP, 100–500 nM free Ca^{2+} (EGTA-buffered), 20 mM Tris pH 7.6 at 37 °C, and 1 mM manganese glutamate to quench the cytosolic fluorescence⁹. This brief treatment with digitonin did not produce leakage of the fluorescence probes or disturb the functional and structural integrity of the endoplasmic reticulum whose reticular structure and ability to respond to $InsP_3$ remained intact after permeabilization.

Optical sectioning. The preparations were imaged by a Nikon Diaphot inverted microscope using a $\times 100$ 1.4 numerical aperture oil-immersion objective and a 100 W mercury vapour epifluorescence illumination source. Images were captured by a thermoelectrically cooled, low-dark-noise (1.3 photoelectrons s^{-1} per pixel at $-36^\circ C$) digital camera with a 16-bit pixel resolution 336×243 charge-coupled-device matrix, and a maximum readout rate of 10^5 pixels s^{-1} (Spectra Source Model 400, Westlake Village, California). The camera was mounted in the photoport of the microscope by using a $\times .9$ relay lens yielding a resolution of 10 pixels μm^{-1} . To increase the sampling rate, we avoided capturing the whole image. Instead, only three-line scans, with an exposure time of 200 ms, were sampled at a rate of 3 scans s^{-1} and sequentially accumulated to form a stack of scans (Fig. 2b). Each scan sampled an area $0.2 \mu m \times 30 \mu m$ across the optical field containing two or three granules or seven to ten segments of endoplasmic reticulum cisterna. Thin optical sections (~ 200 nm) of whole images or of scan stacks were implemented by a no-neighbours deconvolution algorithm⁵. Deconvolution of fluorescently labelled cells depicted the endoplasmic reticulum as a bright reticular network and the mucin granules as bright discrete circular discs (Fig. 2c, d). Thin optical sections allowed us to measure the fluorescent intensity directly as function of time inside or outside the endoplasmic reticulum or secretory granules. We measured fluorescence intensity as the average photoelectron count per pixel in the image area of interest.

Calibration of the optical sectioning technique. The deconvolution program was evaluated by making optical sections of 2 μm fluorescent-coated microspheres (Polysciences, Warrington, Pennsylvania, USA). The algorithm successfully excluded the out-of-focus fluorescence in the images, leaving only fluorescence originating within the shallow depth of field of the objective, and rendering the expected toroidal fluorescent image of the microsphere perimeter (Fig. 2a).

Detection of $[Ca^{2+}]_{ER}$. For single-wavelength fluorescence measurements of $[Ca^{2+}]_{ER}$, ciliated cells were loaded with 500 ng ml^{-1} DiOC₆(3) (Molecular Probes) for 5–8 min and then washed in Hanks' solution. The cells were subsequently equilibrated for 30 min in 10 μM calcium orange-5N-acetoxymethyl ester (Molecular Probes) dissolved in Hanks' solution from a stock in DMSO at 37 °C. The K_d of the Ca^{2+} probes was calculated from direct measurements performed on cells treated with the Ca^{2+} ionophore A23187 by following procedures described elsewhere³⁰. The different excitation wavelengths of DiOC₆(3) and calcium orange-5N (484 and 554 nm respectively) allowed the selection of the image planes within the cell by using DiOC₆(3) without photobleaching calcium orange-5N.

For ratiometric measurements of $[Ca^{2+}]_{ER}$, a separate set of cells (eight batches) were loaded with 20 μM Mag-Indo-1 acetoxymethyl ester ($K_d = 35 \mu M$; Molecular Probes) by using a protocol similar to that for calcium orange-5N. Mag-Indo fluorescence was detected by two photomultiplier tubes (PMTs) (Hamamatsu HC124) mounted on the photoports of a Nikon Multi-image Module (Model 84411). Mag-Indo-1 was excited at 365 nm, and fluorescence emission was collected at 405 and 500 nm by each

of the PMTs. Photon count readings were sampled at 1 kHz by a pulse-counter board connected to a Pentium microprocessor. Free $[Ca^{2+}]$ was calculated by using procedures published elsewhere³⁰.

Ca^{2+}/K^+ ion exchange. Permeabilized ciliated cells or isolated mucin granules were loaded with calcium orange-5N and equilibrated in ATP-free intracellular buffer containing 100 nM free Ca^{2+} (EGTA-buffered), 100 $\mu g\ ml^{-1}$ heparin, and 100 nM apamin^{11,12}. Under these conditions the resting $[Ca^{2+}]_{ER}$ (or $[Ca^{2+}]_G$) remains stable suggesting that the $InsP_3$ receptor channel, the ASK_{Ca} channel and the SERCA pump of the endoplasmic reticulum were rendered inoperative. In addition, 10 μM nigericin, a $K^+ - H^+$ ionophore, was used to equilibrate K^+ and H^+ across the membrane¹⁵. The $[K^+]$ in the intracellular buffer was varied from 1 to 140 mM at pH 5.5 (buffered with 10 mM MES). The ionic strength and osmolality were kept constant by adjusting the concentration of the monovalent organic cation NMG in the intracellular solution.

Received 28 May; accepted 18 August 1998.

- Berridge, M. J. The AM and FM of calcium signalling. *Nature* **386**, 759–760 (1997).
- Koch, G. L. E. The endoplasmic reticulum and calcium storage. *BioEssays* **12**, 527–531 (1990).
- Marszalek, P. E., Farrell, B., Verdugo, P. & Fernandez, J. M. Kinetics of release of serotonin from isolated secretory granules. II. Ion-exchange determines the diffusivity of serotonin. *Biophys. J.* **73**, 1169–1183 (1997).
- Verdugo, P. In *Airway Secretion: Molecular Biophysics of Mucin Secretions* (eds Takishima, T. & Shimura, S.) 101–117 (Marcel Dekker, New York, 1994).
- Monck, J. R., Oberhauser, A. F., Keating, T. J. & Fernandez, J. M. Thin-section ratiometric Ca^{2+} images obtained by optical sectioning of Fura-2 loaded mast cells. *J. Cell Biol.* **116**, 745–759 (1992).
- Breckenridge, L. J. & Almers, W. Final steps in exocytosis observed in a cell with giant secretory granules. *Proc. Natl Acad. Sci. USA* **84**, 1945–1949 (1987).
- Terasaki, M., Song, J., Wong, J. R., Weiss, M. J. & Chen, L. B. Localization of endoplasmic reticulum in living and glutaraldehyde-fixed cells with fluorescent dyes. *Cell* **38**, 101–108 (1984).
- Belan, P. V. et al. A new technique for assessing the microscopic distribution of cellular calcium exit sites. *Pflügers Arch.* **433**, 200–208 (1996).
- Tse, T. W., Tse, A. & Hille, B. Cyclic Ca^{2+} changes in intracellular stores of gonadotropes during gonadotropin-releasing hormone-stimulated Ca^{2+} oscillations. *Proc. Natl Acad. Sci. USA* **91**, 9750–9754 (1994).
- Lyttton, J., Westlin, M. & Hanley, M. R. Thapsigargin inhibits the sarcoplasmic or endoplasmic reticulum Ca-ATPase family of calcium pumps. *J. Biol. Chem.* **266**, 17067–17071 (1991).
- Hehl, S., Golard, A. & Hille, B. Involvement of mitochondria in intracellular calcium sequestration by rat gonadotropes. *Cell Calcium* **20**, 515–524 (1996).
- Ghosh, T. K., Eis, P. S., Mullaney, J. M., Ebert, C. L. & Gill, D. L. Competitive, reversible, and potent antagonism of inositol 1,4,5-trisphosphate-activated calcium release by heparin. *J. Biol. Chem.* **263**, 11075–11079 (1988).
- Latorre, R., Oberhauser, A., Labarca, P. & Alvarez, O. Varieties of calcium-activated potassium channels. *Annu. Rev. Physiol.* **51**, 385–399 (1989).
- Sasaki, S., Nakagaki, I., Kondo, H. & Hori, S. Changes in element concentrations induced by agonist in pig pancreatic acinar cells. *Pflügers Arch.* **432**, 538–545 (1996).
- O'Rourke, F. et al. Ca^{2+} release by inositol (1,4,5)-trisphosphate is blocked by the K^+ -channel blockers apamin and tetraethylammonium ion, and a monoclonal antibody to a 63 kDa membrane protein: reversal of blockade by K^+ ionophores nigericin and valinomycin and purification of the 63 kDa antibody-binding protein. *Biochem. J.* **300**, 673–683 (1994).
- Joseph, S. K. & Williamson, J. R. Characteristics of inositol trisphosphate-mediated Ca^{2+} release from permeabilized hepatocytes. *J. Biol. Chem.* **261**, 14658–14664 (1986).
- Meldolesi, J. & Pozzan, T. The endoplasmic reticulum Ca^{2+} store: a view from the lumen. *Trends Biol. Sci.* **23**, 10–14 (1998).
- Ikemoto, N., Antonui, B., Kang, J. J., Meszaros, L. G. & Rojat, M. Intravesicular calcium transient during calcium release from sarcoplasmic reticulum. *Biochemistry* **30**, 5230–5237 (1991).
- Mitchell, R. D., Simmerman, H. K. B. & Jones, L. R. Ca^{2+} binding effects on protein conformation and protein interactions of canine cardiac calsequestrin. *J. Biol. Chem.* **263**, 1376–1381 (1988).
- Forstner, J. F. & Forstner, G. G. Calcium binding to intestinal goblet mucin. *Biochim. Biophys. Acta* **386**, 283–292 (1975).
- van de Put, F. H. M. M. & Elliott, A. C. The endoplasmic reticulum can act as a functional Ca^{2+} store in all subcellular regions of the pancreatic acinar cell. *J. Biol. Chem.* **272**, 27764–27770 (1997).
- Blondel, O., Bell, G. I. & Seino, S. Inositol 1,4,5-trisphosphate receptors, secretory granules and secretion in endocrine and neuroendocrine cells. *Trends Neurosci.* **18**, 157–161 (1995).
- Gerasimenko, O. V., Gerasimenko, J. V., Belan, P. V. & Petersen, O. H. Inositol trisphosphate and cyclic ADP-ribose-mediated release of Ca^{2+} from single isolated pancreatic zymogen granules. *Cell* **84**, 473–480 (1996).
- Yule, D. I., Ernst, S. A., Ohnishi, H. & Wojcikiewicz, R. J. H. Evidence that zymogen granules are not a physiologically relevant calcium pool. *J. Biol. Chem.* **272**, 9093–9098 (1997).
- Lopez, H. P. & Da Prada, M. Fluorescence microscopical study of 5-hydroxytryptamine storage organelles in mepacrine-incubated blood platelets of beige mice (Chediak-Higashi syndrome). *Experientia* **34**, 663–664 (1978).
- Verdugo, P. Ca^{2+} -dependent hormonal stimulation of ciliary activity. *Nature* **283**, 764–765 (1980).
- Davis, C. W. In *Airway Mucus: Basic Mechanisms and Clinical Perspectives* (eds Rogers, D. F. & Lethem, M. I.) 149–177 (Springer-Verlag, Basel, 1997).
- Finch, E. A., Turner, T. J. & Goldin, S. M. Calcium as a coagonist of inositol 1,4,5-trisphosphate-induced calcium release. *Science* **252**, 443–446 (1991).
- Aitken, M. L., Villalon, M., Pier, M., Verdugo, P. & Nameroff, M. Characterization of a marker tracheal basal cells. *Exp. Lung Res.* **21**, 1–16 (1995).
- Kao, J. P. Y. In *A Practical Guide to the Study of Calcium in Living Cells* (ed. Nuccitelli, R.) 155–181 (Academic, San Diego, 1994).

Acknowledgements. We thank M. Pier for assistance in cell culture, and P. Detwiler, J. M. Fernandez, R. Hart, B. Hille, T. Hinds, A. Gordon, G. Pollack and J. L. Vergara for suggestions and the critical review of the manuscript. This work was funded by grants from USA NSF, the USA DOE and the CONRAD Program of Mellon Foundation. W.-C.C. and T.N. were supported by scholarships from the Center for Nanotechnology and the MSTP Program from the University of Washington, respectively.

Correspondence and requests for materials should be addressed to P.V. (e-mail: verdugo@bioeng.washington.edu).

Structure of a glutamate-receptor ligand-binding core in complex with kainate

Neali Armstrong^{*†}, Yu Sun^{*†}, Guo-Qiang Chen^{*} & Eric Gouaux^{*}

^{*} Department of Biochemistry and Molecular Biophysics, Columbia University, 650 W. 168th Street, New York, New York 10032, USA

[†] These authors contributed equally to this work

Ionotropic glutamate receptors (iGluRs) mediate excitatory synaptic transmission in vertebrates and invertebrates through ligand-induced opening of transmembrane ion channels. iGluRs are segregated into three subtypes according to their sensitivity to the agonists AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid), kainate (a structural analogue of glutamate) or NMDA (*N*-methyl-D-aspartate) (Fig. 1). iGluRs are important in the development and function of the nervous system, are essential in memory and learning, and are either implicated in or have causal roles in dysfunctions ranging from Alzheimer's, Parkinson's and Huntington's diseases, schizophrenia, epilepsy and Rasmussen's encephalitis to stroke^{1,2}. Development of iGluR agonists and antagonists has been hampered by a lack of high-resolution structural information. Here we describe the crystal structure of an iGluR ligand-binding region in a complex with the neurotoxin (agonist) kainate. The bilobed structure shows the determinants of receptor-agonist interactions and how ligand-binding specificity and affinity are altered by remote residues and the redox state of the conserved disulphide bond. The structure indicates mechanisms for allosteric effector action and for ligand-induced channel gating. The information provided by this structure will be essential in designing new ligands.

The iGluR ligand-binding core is formed by polypeptide segments S1 and S2 (refs 3, 4). S1 includes ~150 residues amino-

terminal to the first transmembrane segment and S2 comprises the residues between membrane-associated segments 3 and 4 (Fig. 2). Definition of the S1S2 region as an autonomous ligand-binding core for all iGluRs is strengthened by site-directed mutagenesis and chimaeric and S1-S2 fusion experiments⁵. A construct in which the S1 and S2 regions of GluR4 are coupled by a 13-residue linker has pharmacological properties that are similar to those of the wild-type membrane-bound receptor⁴. We have previously developed methods to obtain large amounts of iGluR S1S2 protein from *Escherichia coli* and have generated new S1S2 constructs that form well-ordered crystals⁶.

We solved the structure of the rat GluR2 S1S2 'flop' isoform bound to the non-desensitizing agonist kainate⁷ by multiwavelength anomalous diffraction (MAD)⁸ from crystals of the selenomethionine derivative. Of the 279 residues in the crystallized protein, there is corresponding electron density for 250 residues. The N-terminal 15 residues, 3 residues in loop 1, 9 amino acids in the S1-S2 linker region and the Gly-Ser dipeptide at the end of S2 are disordered in the crystal. Crystallographic statistics and views of the structure are shown in Table 1 and Figs 2 and 3.

The S1S2 structure has two α/β domains arranged in the shape of a kidney. It is 57 Å in length and has cross-sectional dimensions of 43 Å and 35 Å. The N terminus of the S1S2 construct is located in domain 1 and the carboxy terminus is at the end of helix K and is disulphide-bonded to domain 2. Domains 1 and 2 are composed mainly of polypeptide segments S1 and S2, respectively. However, S1 crosses over S2 and ends in domain 2, while the C terminus of S2 constitutes part of domain 1. Kainate is nestled between domains 1 and 2 and participates in multiple, primarily polar, contacts with residues from both domains. There are few, but significant, sequence relationships between GluR2 S1S2 and bacterial periplasmic-ligand binding proteins⁹; consistent with these relationships, the structure of the S1S2 construct is strikingly similar to that of the *E. coli* glutamine-binding protein (QBP)¹⁰.

Kainate binds in the interdomain crevice at the N termini of four α -helices; it induces domain closure and resistance to proteolysis⁶. The 2-carboxyl group forms essential interactions with the guanidinium group of Arg 485 and the main-chain NH group of Thr 480 of GluR2; mutation of Arg 485 to Lys or Ser at the corresponding sites in the homologous GluR4, NMDAR1 and NMDAR2B receptors and in chick kainate-binding protein (KBP) abolishes ligand-binding activity or agonist-induced opening of the ion channel¹¹⁻¹³. Exploiting the macro dipole of helix F, the carboxyl methyl group forms hydrogen bonds with the NH groups of Ser 654 and Thr 655, the hydroxyl group of Thr 655, and two water molecules tethered to the peptide carbonyl oxygen of Ser 652 and the peptide nitrogen of Glu 705. Mutagenesis of Thr 268 to Ala in chick KBP (Thr 655 in GluR2) abolishes binding to ³H-kainate¹³. Tyr 450 acts like a lid on top of the pyrrolidine ring and isopropenyl group of kainate, while the amino group is bound by the essential carboxylate of Glu 705, the carbonyl oxygen of Pro 478, and the sidechain hydroxyl of Thr 480. In NMDA receptors, an aspartate replaces the equivalent of Glu 705 in GluR2, perhaps to accommodate the 'b' stereocentre. The electrostatic potential of the ligand-binding site is highly negative and the net charge on kainate, AMPA and glutamate is -1. These apparently destabilizing electrostatic effects are compensated for by specific hydrogen bonds and salt links, and by helical dipole effects, as the N termini of four helices are located in the ligand site.

The isopropenyl group is partially exposed to solvent and projects into the interdomain crevice from underneath the Tyr 450 lid. The dearth of interactions with the isopropenyl group may be one reason why kainate binds relatively weakly to GluR2. Furthermore, the orientation of the isopropenyl group explains how ligands such as domoate, which have longer functional groups located at the 4 position of the pyrrolidine ring than does kainate, bind to the receptor. Four sequestered water molecules are next to the

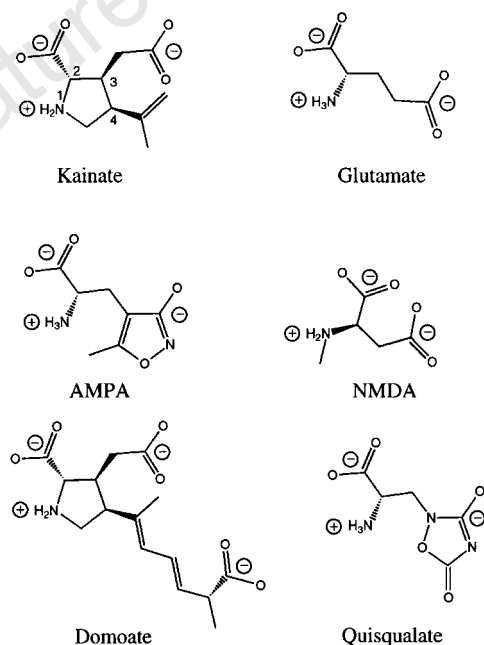


Figure 1 Chemical structures of some iGluR agonists. The numbering of substituents on the pyrrolidine ring of kainate is as follows: nitrogen, 1; carboxylate, 2; carboxyl methyl, 3; and isopropenyl, 4.

isopropenyl and carboxyl methyl groups.

Ligand-binding affinity and specificity are altered by multidendate ligand–receptor interactions, which, in turn, induce and control domain separation. We suggest that binding of different agonists may result in variable degrees of domain closure. Binding of AMPA and glutamate to GluR2 S1S2 may allow more domain

closure than kainate. This is because AMPA and glutamate do not have the pyrrolidine ring and isopropenyl groups that together act like an interdomain wedge against the Tyr 450 lid. Fitting a model of AMPA onto the kainate structure shows that AMPA can supplant all of kainate's interactions, can accept an additional hydrogen bond through the isoxazole ring, and can allow greater domain closure.

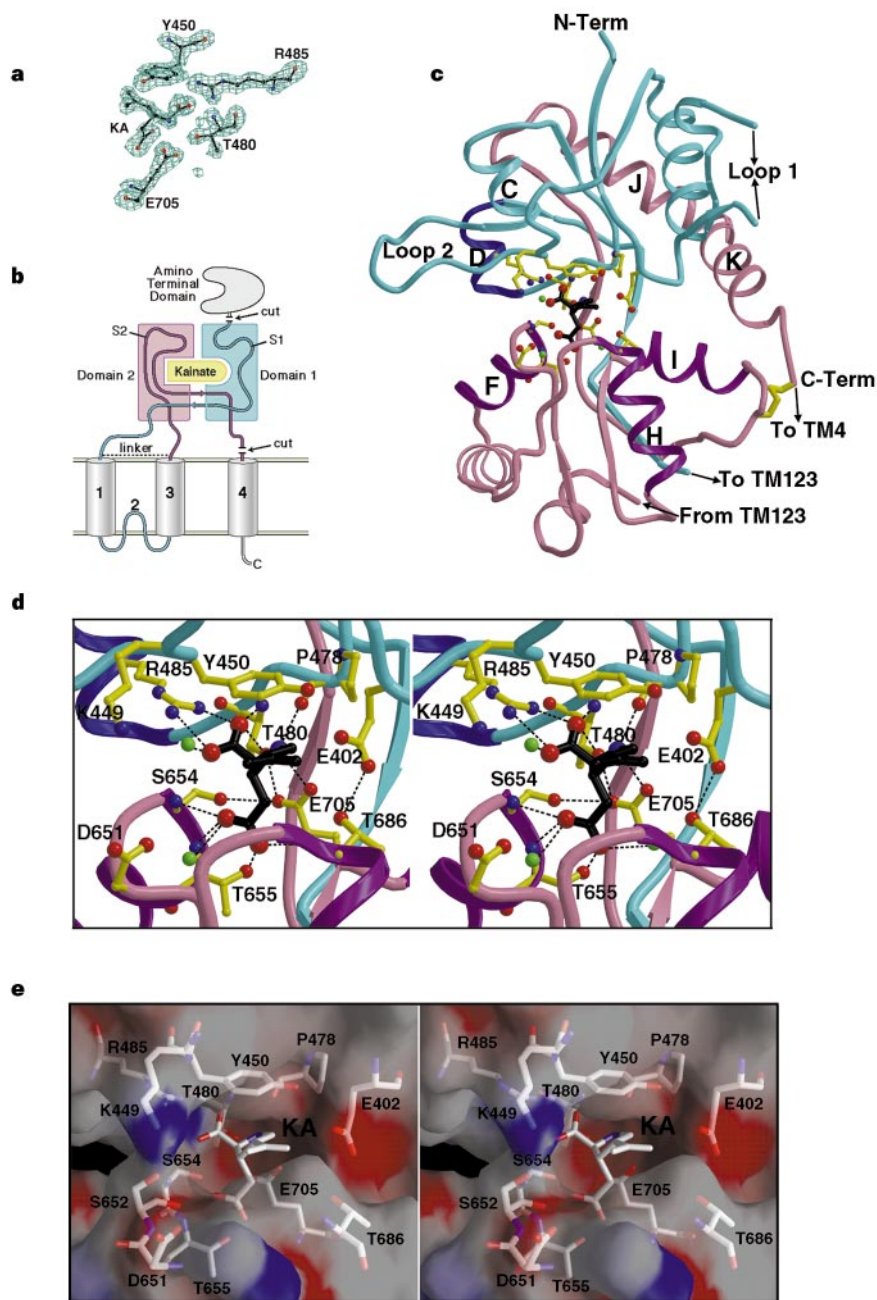


Figure 2 Structure of the complex of GluR2 S1S2 and kainate. **a**, $A F_o - F_c$ omit electron-density map of kainate and four selected ligand-binding residues, calculated using coefficients between 6.0 and 1.9 Å resolution and phases from the refined model. The map is contoured at 2.5σ . **b**, Domain structure of iGluRs, showing the S1 and S2 segments in turquoise and pink, respectively. 'Cut' and 'linker' denote the edges of the S1S2 construct. **c**, Ribbon representation of the S1S2 structure, with S1 and S2 coloured as in **b**. The N termini of helices D, F, H and I each contain a residue that either interacts with kainate (black ball-and-stick representation) or participates in interdomain contacts: Arg 485 projects into the kainate-binding site from helix D (dark purple); the backbone amides of Ser 654 and Thr 655 on helix F (light purple) interact with the carboxyl methyl group of

kainate; and helices H and I (light purple) contain residues Thr 686 and Glu 705. The disulphide bond between Cys 718 (helix I) and Cys 773 is shown in yellow. The C terminus of S1 is marked 'To TM123' and the N terminus of S2 is labelled 'From TM123'. Loops 1 and 2 protrude from the protein core by at least 10 Å and may be involved in domain–domain contacts. TM, transmembrane domain. **d**, Stereo view of the S1S2 region that encompasses kainate (black), showing interacting residues and residues that make interdomain contacts. The dashed lines indicate potential hydrogen bonds and ionic interactions between residues $<3.2 \text{ Å}$ apart. Water molecules are shown as green balls. **e**, Stereo view of the molecular surface, coloured by electrostatic potential; red, negative; blue, positive. Two regions of continuous molecular surface are formed on either side of kainate (KA).

Furthermore, the mode of ligand–receptor interaction seen in the complex of GluR2 S1S2 and kainate explains how glycine, which lacks an equivalent of kainate's carboxyl methyl group, can function as an NMDA-receptor agonist¹⁴: the amino and carboxyl groups bridge both domains, and promote domain closure and specific interdomain interactions.

Control of ligand affinity and specificity and of the kinetic properties of the ion channels also occurs through allosteric interactions and involves residue pairs on domain 1 and domain 2 that flank kainate (Glu 402–Thr 686 and Lys 449–Asp 651, Ser 652; Fig. 2). The residue at position 686 of GluR2 in the kainate receptors GluR5 (Ser) and GluR6 (Asn) is important in determining sensitivity to AMPA and the rate of deactivation by domoate¹⁵. In the sulphate-binding protein, disruption of similar interdomain interactions significantly increases the K_d value and k_{off} rate for sulphate¹⁶. As seen in GluR2 S1S2, Glu 402 forms a hydrogen bond with the hydroxyl group of Thr 686. Although Lys 449 (which is in domain 1) does not contact residues on domain 2 directly, it is only ~4 Å from Asp 651 and Ser 652. If further domain closure were to occur upon glutamate or AMPA binding, Lys 449 could participate in an interdomain salt link or hydrogen bond. In fact, when the equivalent lysine in GluR1 is mutated to glutamine, the effector concentrations for half-maximal response (EC_{50} values) for glutamate and AMPA increase; there is no effect on the EC_{50} for kainate¹⁷.

A comparison of the kainate–S1S2 structure and the QBP–glutamine structure (Fig. 3)¹⁰ supports the theory that the GluR2 S1S2 domains may undergo further domain closure. When domain 1 of S1S2 is superimposed on domain 1 of the closed form of QBP, a rotation of 21° around an axis located between the domains is needed to superimpose domains 2. When the open form of QBP is used as the template, a rotation of 32° is required to fit domain 2 of S1S2 on domain 2 of QBP. Compared with the open and closed forms of QBP, the GluR2 S1S2 complex containing kainate is intermediate with respect to domain closure. To determine whether further domain closure could occur in the presence of different ligands, such as AMPA, we independently fit domains 1 and 2 of S1S2 to the closed form of QBP by creating two interdomain breaks in the polypeptide chain. In the resulting S1S2 model there were only a few steric clashes, involving Lys 449 and Ser 652, and Glu 402 and Met 708. Thus, further domain closure is possible if a few residues and associated elements of the main chain undergo conformational changes.

Redox modulation of NMDA-receptor activity, in which disulphide-reducing agents increase NMDA-gated currents and oxidizing agents decrease the currents¹⁸, may be understood in terms of the conserved interdomain disulphide bond between Cys 718 and Cys 773. Redox control of the receptor may occur by making domain closure more difficult (in the oxidized state) or easier (in the reduced state). By analogy with QBP, in domain-closed forms of the receptor the distance between the cysteine residues is larger than in more open forms and thus cleavage of the disulphide bond by reduction or mutagenesis increases ligand-binding affinity or ligand-induced current by allowing domain closure¹⁹. Domain movement in the oxidized state probably also involves contraction and extension of the polypeptide chain at the end of helix K.

Sequence alignment of regions of iGluRs that include domains corresponding to the S1S2 construct of GluR2 shows that all iGluRs have a common core, indicates the site of allosteric effector action, and maps the location of the flip/flop residues encoded by separate exons⁷. All seven residues that make direct interactions with kainate in the GluR2 S1S2 complex are identical or conservatively substituted among the members of the four iGluR subfamilies, which include the kainate-binding proteins, even though the overall degree of sequence identity between iGluRs is <10% (Fig. 4). All four flip/flop sites in this GluR2 S1S2 construct, in addition to the RNA-editing site, are located on the solvent-exposed surface of helices J and K. The flip (Ser) and flop (Asn) variants at position 754 of GluR2 have differential allosteric responses to cyclothiazide. Kainate receptors, which contain a Gln at this position, are insensitive to cyclothiazide. Mutation of Ser 750 to Gln in the AMPA receptor GluR1 (position 754 in GluR2) abolishes modulation by cyclothiazide, whereas introduction of Ser at the equivalent site in the kainate receptor GluR6 confers sensitivity to cyclothiazide²⁰. The clustering of Asn 754, Leu 748 and Leu 751 indicates that this

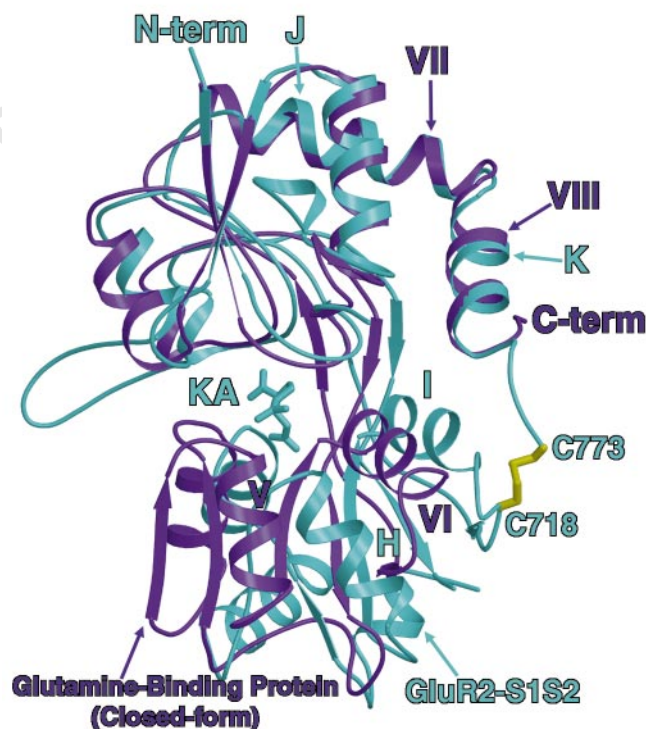


Figure 3 Superposition of the complex of GluR2 S1S2 and kainate (KA) with the closed form of QBP¹⁰. Helices I, VII and VIII from QBP were fit to helices B, J and K, respectively, of GluR2 S1S2; of these, helices VII, VIII, J and K are labelled. Following this superposition, the root mean square (r.m.s.) deviations of α -carbon positions is 1.0 Å for the atoms fit, 1.6 Å for domain 1, 8.5 Å for domain 2 and 5.5 Å for both domains. Subsequent fitting of domains 2 from QBP and GluR2 S1S2 gives r.m.s. deviations of α -carbon positions of 2.1 Å for domain 2, 6.6 Å for domain 1 and 5.2 Å overall.

Table 1 Refinement statistics

Resolution (Å)	Reflections (<i>N</i>)	<i>R</i> * (%)	<i>R</i> _{free} † (%)	Average <i>B</i> -value			r.m.s. deviations			
				Overall	Main	Ligand	Bonds	Angles	<i>B</i> -values	
									Bonds	Angles
6.0–1.9	19,483	21.6	27.3	18.86	17.53	11.07	0.011 Å	1.483°	2.5	3.5

The model included 250 residues, 175 water molecules, 1 kainate molecule and 2,068 atoms.

* R -factor = $(\sum ||F_o| - |F_c||) / \sum |F_o|$, where F_o and F_c denote observed and calculated structure factors, respectively, and summation extends over all observed reflections.

† 8% of the structure factors between 6.0 and 1.9 Å were set aside to calculate the R -free factor.

surface is the site of effector action and that, because of its hydrophobic character, this surface is probably involved in intra-subunit or intersubunit contacts; in other words, cyclothiazide, aniracetam and thiocyanate²¹ may function by perturbing interfaces that are >25 Å from the ligand-binding site. The agonist-like effects of autoantibodies implicated in Rasmussen's encephalitis² also act at a substantial distance from the ligand-binding site: on the basis of the S1S2 structure, the epitope(s) are located just upstream of the start of S1, near helix J.

This is, to our knowledge, the first high-resolution structure of an iGluR ligand-binding core, and it revealed the following unexpected facts. First, the ligand-binding site of GluR2 has a negative electrostatic potential. Second, the carboxyl methyl group of kainate is neutralized by hydrogen bonds from the N terminus of helix F. Third, the binding of kainate to GluR2 S1S2 induces an intermediate degree of domain closure relative to QBP's open and closed

conformations; this degree of closure corresponds to the open, non-desensitized state of the intact receptor. Fourth, some of the residues that mediate agonist specificity and channel dynamics form interdomain contacts. Finally, the exposed face of helix J is probably a site of binding of allosteric effectors and of domain-domain contacts.

Methods

Crystallization and data collection. We expressed, purified, characterized and crystallized the GluR2 S1S2 construct as described⁶. The space group and unit-cell dimensions for the S1S2-kainate co-crystals are P_2 , and $a = 42.22$ Å, $b = 63.07$ Å, $c = 46.32$ Å and $\beta = 92.99^\circ$. The methionine and selenomethionine (SeMet) crystals were isomorphous. Diffraction data sets were collected at the NSLS X4a and CHESS F-2 beamlines from selenomethionine crystals that were flash-cooled in liquid nitrogen and maintained in a stream of nitrogen gas at -150°C . Before cooling, the crystals were soaked in the crystallization reservoir

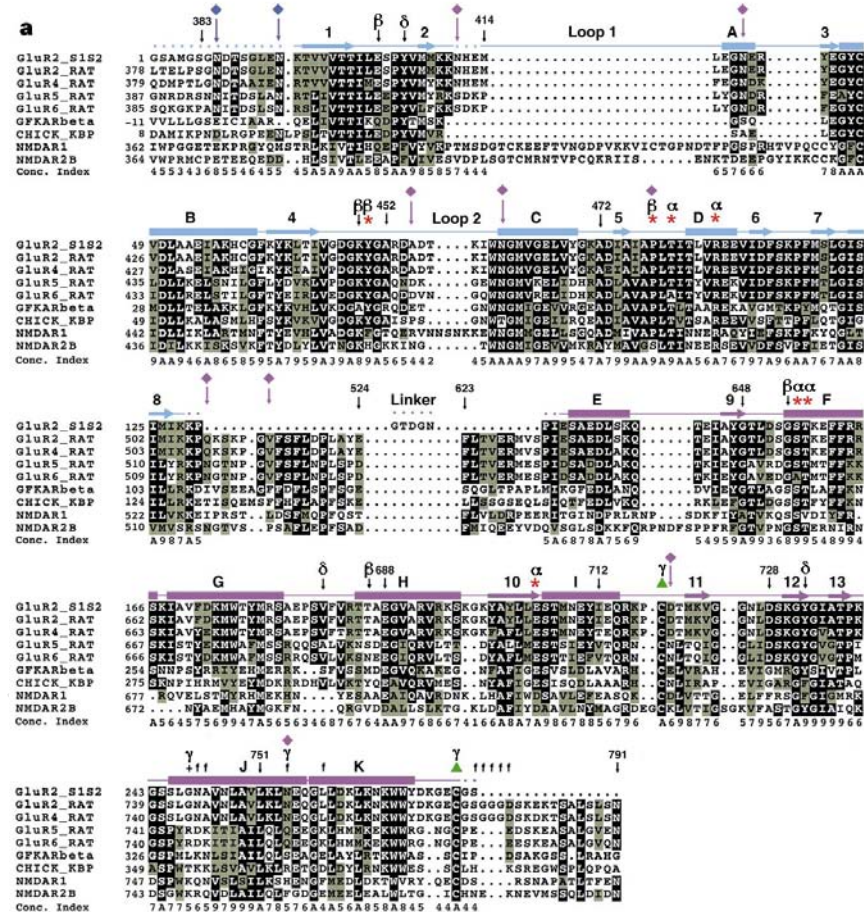


Figure 4 Sequence identity between iGluRs. **a**, Multiple sequence alignment of selected AMPA, kainate and NMDA receptors and kainate-binding proteins. Residues are defined as not conserved (conservation index of 0) through to moderately well-conserved (index of 8) and highly conserved (index of A). α -Helices (rectangles) and β -strands (arrows) are drawn above the sequences, with colouring as in Fig. 2; dotted lines indicate regions for which there is no electron density available for the main chain. Ligand-binding residues are indicated by red stars. The disulphide-bond-forming cysteines are marked by green triangles. Flip/flop residues of GluR2 are delineated by 'f' and the RNA-editing site by '+'. Glycosylation sites in GluR1 are labelled with diamonds, which are magenta for engineered sites and blue for natural sites²⁰. Functionally important residues are defined as follows: α , residues interacting with all agonists; β , residues that interact with specific ligands; γ , residues involved in the regulation of ligand binding and the kinetic properties of the channel; δ ,

residues that do not interact directly with the ligand but may be important for the structure of the binding pocket. Sequences are numbered according to the predicted mature polypeptide. **b**, The degree of conservation (0, white, to A, black) is mapped onto each residue position of GluR2 S1S2. The four flip/flop sites (Asn 744, Ala 745, Asn 754 and Leu 758; shown in magenta) and the Arg/Gly RNA-editing site (Gly 743) are located on helices J and K. Residue Asn 754 (flip, Gly; flop, Asn) is involved in the allosteric responses to cyclothiazide²⁰. The Arg/Gly RNA-editing site (Gly 743) influences the rate of recovery from desensitization²⁹. **c**, Solvent-accessible surface of GluR2 S1S2, in the same orientation as in **b**. Hydrophobic residues (Ala, Val, Leu, Ile, Met, Pro, Phe, Tyr and Trp) are purple; residues with a conservation index of 9 or A are green; exposed and conserved hydrophobic residues are black. A cylindrical depression, ~ 3.5 Å in diameter and 4.0 Å deep, is located at the Arg/Gly RNA-editing site.

solution supplemented with ~20% (w/v) PEG 400. After using the SeMet crystals to obtain a fluorescence curve as a function of the incident X-ray energy, we collected data sets at 12659.0 eV, 12663.5 eV, 12800.0 eV and 12630.0 eV using inverse-beam geometry. The data sets from the R-Axis4 detector at NSLS X4a and the Quantum 4 detector at CHESS were processed using DENZO and SCALEPACK²².

Structure determination. Determination and refinement of the selenium sites and initial phase calculations were done using the program SOLVE (from T. Terwilliger; see <http://www.solve.lanl.gov> and references therein). Density modification that included solvent flattening and histogram matching, using the program DM²³, improved the quality of the electron-density maps calculated using either the NSLS or the CHESS data. However, the maps derived from the NSLS or CHESS data were mediocre and could not be used to trace more than a small fraction of the polypeptide chain. Substantial improvement of the electron-density maps resulted from combination of the NSLS and CHESS phases using the program SIGMA²⁴ and use of $|F_o|$ derived from a selenomethionine data set collected at home (2.3 Å resolution, CuK α X-rays) and MAD-combined phases. Phase combination in reciprocal space and map averaging in real space produced similar results. Using experimental maps calculated at 2.5 and 2.3 Å resolution, we built a structure comprising 86% of the 279 residues by using the program O (ref. 25). After the first round of Powell minimization and *B*-factor refinement using XPLOR²⁶, the conventional *R*-value was 0.322 and the *R*-free value was 0.389. At this point, we used native data measured to 1.9 Å resolution for the remainder of the refinement and model building until the *R* values converged, beginning at 2.3 Å resolution and gradually including higher-resolution terms. We then built kainate into an 'omit' map calculated using phases from the refined model and $F_o - F_c$ coefficients, and carried out another round of Powell minimization and *B*-factor refinement. Unless otherwise noted, CCP4 programs²⁷ were used for crystallographic calculations. There is either no or weak electron density for the following side chains and they have therefore been omitted from the structure and replaced by Ala: Lys 393, Lys 410, Glu 416, Lys 434, Asp 456, Leu 483, Lys 493, Glu 634, Ser 635, Glu 637, Lys 641, Gln 642, Thr 643, Arg 661, Lys 663, Glu 678, Lys 695, Lys 697 and Ser 741.

Received 14 August; accepted 18 September 1998.

- Hollmann, M. & Heinemann, S. Cloned glutamate receptors. *Annu. Rev. Neurosci.* **17**, 31–108 (1994).
- Rogers, S. W. *et al.* Autoantibodies to glutamate receptor GluR3 in Rasmussen's encephalitis. *Science* **265**, 648–651 (1994).
- Stern-Bach, Y. *et al.* Agonist selectivity of glutamate receptors is specified by two domains structurally related to bacterial amino acid-binding proteins. *Neuron* **13**, 1345–1357 (1994).
- Kuusinen, A., Arvola, M. & Keinänen, K. Molecular dissection of the agonist binding site of an AMPA receptor. *EMBO J.* **14**, 6327–6332 (1995).
- Paas, Y. The macro- and microarchitectures of the ligand-binding domain of glutamate receptors. *Trends Neurosci.* **21**, 117–125 (1998).
- Chen, G.-Q., Sun, Y., Jin, R. & Gouaux, E. Probing the ligand binding domain of the GluR2 receptor by proteolysis and deletion mutagenesis defines domain boundaries and yields a crystallizable construct. *Protein Sci.* (in the press).
- Sommer, B. *et al.* Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. *Science* **249**, 1580–1585 (1990).
- Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
- Nakanishi, N., Schneider, N. A. & Axel, R. A family of glutamate receptor genes: evidence for the formation of heteromultimeric receptors with distinct channel properties. *Neuron* **5**, 569–581 (1990).
- Sun, Y.-J., Rose, J., Wang, B.-C. & Hsiao, C.-D. The structure of glutamine-binding protein complexed with glutamine at 1.94 Å resolution: comparisons with other amino acid binding proteins. *J. Mol. Biol.* **278**, 219–229 (1998).
- Keinänen, K., Arvola, M., Kuusinen, A. & Johnson, M. Ligand recognition in glutamate receptors: insights from mutagenesis of the soluble α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)-binding domain of glutamate receptor type D (GluR-D). *Biochem. Soc. Trans.* **25**, 835–838 (1997).
- Laube, B., Hirai, H., Sturgess, M., Betz, H. & Kuhse, J. Molecular determinants of agonist discrimination by NMDA receptor subunits: analysis of the glutamate binding site on the NR2B subunit. *Neuron* **18**, 493–503 (1997).
- Paas, Y., Eisenstein, M., Medevielle, F., Teichberg, V. I. & Devillers-Thiery, A. Identification of the amino acid subsets accounting for the ligand binding specificity of a glutamate receptor. *Neuron* **17**, 979–990 (1996).
- Hirai, H., Kirsch, J., Laube, B., Betz, H. & Kuhse, J. The glycine binding site of the *N*-methyl-D-aspartate receptor subunit NR1: identification of novel determinants of co-agonist potentiation in the extracellular M3–M4 loop region. *Proc. Natl Acad. Sci. USA* **93**, 6031–6036 (1996).
- Swanson, G. T., Gereau, R. W. IV, Green, T. & Heinemann, S. F. Identification of amino acid residues that control functional behavior in GluR5 and GluR6 kainate receptors. *Neuron* **19**, 913–926 (1997).
- Jacobson, B. L., He, J. J., Lemon, D. D. & Quirocho, F. A. Interdomain salt bridges modulate ligand-induced domain motion of the sulfate receptor protein for active transport. *J. Mol. Biol.* **223**, 27–30 (1992).
- Li, F., Owens, N. & Verdoorn, T. A. Functional effects of mutations in the putative agonist binding region of recombinant α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors. *Mol. Pharmacol.* **47**, 148–154 (1995).
- Aizenman, E., Lipton, S. A. & Loring, R. H. Selective modulation of NMDA responses by reduction and oxidation. *Neuron* **2**, 1257–1263 (1989).

- Sutcliffe, M. J., Wo, Z. G. & Oswald, R. E. Three-dimensional models of non-NMDA glutamate receptors. *Biophys. J.* **70**, 1575–1589 (1996).
- Partin, K. M., Bowie, D. & Mayer, M. L. Structural determinants of allosteric regulation in alternatively spliced AMPA receptors. *Neuron* **14**, 833–843 (1995).
- Partin, K. M., Fleck, M. W. & Mayer, M. L. AMPA receptor flip/flop mutants affecting deactivation, desensitization, and modulation by cyclothiazide, aniracetam, and thiocyanate. *J. Neurosci.* **16**, 6634–6647 (1996).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Enzymol.* **276**, 307–326 (1997).
- Cowan, K. D. & Main, P. Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr. D* **42**, 43–48 (1996).
- Read, R. J. Model phases: probabilities and bias. *Methods Enzymol.* **277**, 110–128 (1997).
- Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
- Brünger, A. T. *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR 1–382* (Yale Univ. Press, New Haven, 1992).
- Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Hollmann, M., Maron, C. & Heinemann, S. N-Glycosylation site tagging suggests a three transmembrane domain topology for the glutamate receptor GluR1. *Neuron* **13**, 1331–1343 (1994).
- Lomeli, H. *et al.* Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **266**, 1709–1713 (1994).

Acknowledgements. We thank C. Lima and other members of the Hendrickson laboratory for assistance; J. Lidestri for maintenance of the X-ray laboratory at Columbia University; A. Karlin, J. Javitch and M. Akabas for advice; S. Heinemann and M. Hartley for glutamate-receptor clones; C. Ogata for help with MAD data collection at NSLS X4a; T. Terwilliger for advice on the use of Solve; and W. Minor and Z. Otwinowski for a new version of the HKL suite of programs. Extra MAD data were collected at the Cornell High Energy Synchrotron Source (CHESS), which is supported by the NSF, using the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by an award from the NIH. N.A. was supported by a NIH Ophthalmology training grant. E.G. is an NSF Young Investigator and the recipient of an Alfred P. Sloan Research Fellowship.

Correspondence and requests for materials should be addressed to E.G. (e-mail: jeg52@columbia.edu). The GluR2 S1S2 coordinates have been deposited with the Protein Data Bank and have accession code 1gr2.

Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex

Jeffrey K. Tong*, Christian A. Hassig*, Gavin R. Schnitzler†, Robert E. Kingston† & Stuart L. Schreiber*

* Howard Hughes Medical Institute, Departments of Chemistry and Chemical Biology, Molecular and Cellular Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

† Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

The dynamic assembly and remodelling of eukaryotic chromosomes facilitate fundamental cellular processes such as DNA replication and gene transcription. The repeating unit of eukaryotic chromosomes is the nucleosome core, consisting of DNA wound about a defined octamer of histone proteins¹. Two enzymatic processes that regulate transcription by targeting elements of the nucleosome include ATP-dependent nucleosome remodelling and reversible histone acetylation^{2,3}. The histone deacetylases, however, are unable to deacetylate oligonucleosomal histones *in vitro*⁴. The protein complexes that mediate ATP-dependent nucleosome remodelling and histone acetylation/deacetylation in the regulation of transcription were considered to be different, although it has recently been suggested that these activities might be coupled⁵. We report here the identification and functional characterization of a novel ATP-dependent nucleosome remodelling activity that is part of an endogenous human histone deacetylase complex. This activity is derived from the CHD3 and CHD4 proteins which contain helicase/ATPase domains found in SWI2-related chromatin remodelling factors, and facilitates the deacetylation of oligonucleosomal histones *in vitro*. We refer to this complex as the nucleosome remodelling and deacetylating (NRD) complex. Our results establish a physical and functional link between the distinct chromatin-modifying activities of histone deacetylases and nucleosome remodelling proteins.

We purified an endogenous histone deacetylase complex, HDAC2 complex, from HeLa cell extract by immunoaffinity chromatography. The antibody-bound complex was specifically eluted by its peptide antigen, separated by SDS-PAGE and visualized by silver staining (Fig. 1). As expected, the HDAC2-associated proteins RbAp48 and HDAC1 co-purified with HDAC2. In addition, three

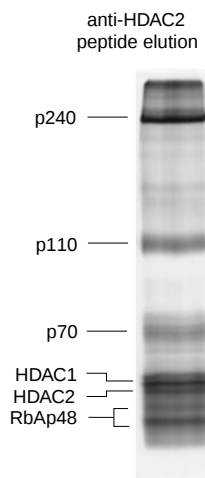


Figure 1 Silver-stained SDS-polyacrylamide gel of the HDAC2 complex. The HDAC2-associated polypeptides p70, p110 and p240 were immunoaffinity-purified from HeLa cell extract as described in Methods.

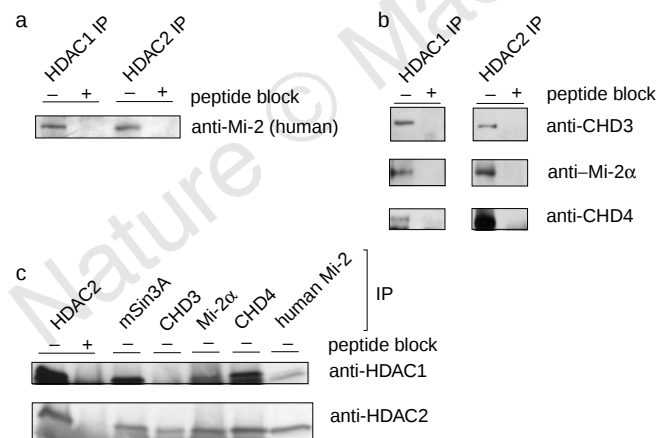


Figure 2 CHD3 and CHD4 are associated with HDAC1 and HDAC2 *in vivo* in the NRD complex. **a**, Jurkat and HeLa cell extracts were immunoprecipitated using anti-HDAC1 or anti-HDAC2 antisera, respectively. Bound proteins were separated by SDS-PAGE and analysed by western blotting with human Mi-2-positive serum. As negative controls, parallel immunoprecipitations were done using antiserum preincubated with the peptide antigen. **b**, As **a**, except that western blots were prepared with two different anti-CHD3 rabbit sera (anti-CHD3 and anti-Mi-2α) as well as an anti-CHD4 rabbit serum. **c**, Immunoprecipitation from HeLa extract using rabbit antisera against CHD3 and CHD4, and human Mi-2-positive dermatomyositis patient serum. mSin3A and HDAC2 immunoprecipitations were tested in parallel as positive controls for HDAC1 and HDAC2. Antibody-bound proteins were separated by SDS-PAGE and analysed by western blotting for HDAC2 (bottom). The HDAC2 blot was stripped and reprobed for HDAC1 (top). HDAC1 corresponds to the top band of the doublet seen in the CHD4 and Mi-2 lanes.

previously uncharacterized HDAC-associated polypeptides of apparent relative molecular masses of 70K, 110K and 240K (referred to as p70, p110 and p240, respectively) were identified. The purified proteins were excised from the gel for sequence analysis by peptide mass spectrometry. These three uncharacterized polypeptides are also common to an HDAC1 immune complex prepared from human Jurkat T cells (data not shown). Peptides obtained from p110 are identical in sequence to a region of the hypothetical protein KIAA0601 (GenBank accession number, AB011173). This protein shares a region of homology with a polyamine oxidase, suggesting a potential role for the NRD complex in redox reactions of straight-chain aminoalkane structures, such as those that occur in lysine side chains and in the chromatin cofactors spermine and spermidine⁶. Here we focus on the characterization of p240. Sequences from two peptides derived from p240 correspond to sequences found in the related human proteins CHD3 and CHD4; peptide 1 (FSWAQGTDTILADEMGLGK) is an identical match to a sequence found in both CHD3 and CHD4; peptide 2 (HDYWLLAGIINHGYAR) is unique to CHD4. CHD3 and CHD4 have previously been identified as self-antigens recognized by sera from a subset of patients with the autoimmune connective-tissue disease Mi-2 dermatomyositis^{7,8}.

We investigated whether CHD3 and CHD4 are present in both Jurkat HDAC1 and HeLa HDAC2 complexes by co-immunoprecipitation and western blotting analysis. HDAC1 and HDAC2 complexes were immunoprecipitated from Jurkat and HeLa extracts, respectively, and then western-blotted with Mi 2-positive dermatomyositis patient serum, which recognizes both CHD3 and CHD4 (Fig. 2a). The serum specifically reacts with p240 in both the HDAC1 and HDAC2 complexes and no crossreacting band was seen when HDAC immunoprecipitation was pre-blocked with the peptide antigen (Fig. 2a). To confirm the presence of CHD3 and CHD4 in the complexes, HDAC immunoprecipitates were probed with two different rabbit antisera raised against CHD3 (anti-CHD3 and anti-Mi-2α) and one other rabbit antiserum raised against CHD4. These antibodies were generated against regions of CHD3

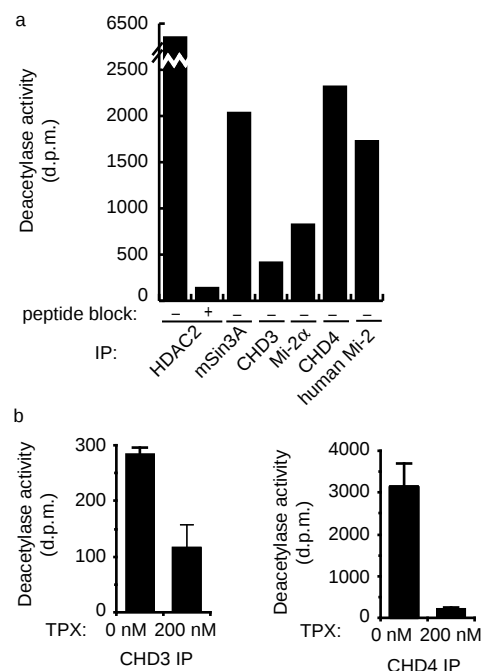


Figure 3 CHD3 and CHD4 complexes contain histone deacetylase activity. **a**, Each immunoprecipitate from Fig. 2c was tested for deacetylase activity. **b**, CHD3 and CHD4 were immunoprecipitated from HeLa cell extract and the purified complex was preincubated with or without the HDAC inhibitor trapoxin (TPX) before assaying for histone deacetylase activity. Background was ~100 d.p.m.

and CHD4 that are unique to each protein. All anti-CHD3 and anti-CHD4 antisera reacted with p240 in an HDAC-specific manner (Fig. 2b). Reciprocal immunoprecipitation using anti-CHD3 or anti-CHD4 antisera, or Mi-2 human serum, were positive for HDAC1 and HDAC2 as well (Fig. 2c).

To determine whether HDACs associated with the CHD proteins in the NRD complex were enzymatically active, we tested the CHD3 and CHD4 immunoprecipitates for histone deacetylase activity. *In vitro* activity was measured by quantifying the release of radiolabelled acetyl groups from purified hyperacetylated HeLa histones. We found that CHD3 and CHD4 immunoprecipitates from HeLa cell extract had histone deacetylase activity (Fig. 3a), and that treatment of these immunoprecipitates with the HDAC inhibitor trapoxin reduces histone deacetylase activity to background levels (Fig. 3b).

The presence of CHD3 and CHD4 in the NRD complex prompted us to investigate their possible function in chromatin regulation. CHD3 and CHD4 are highly related proteins that contain a number of putative functional domains, including PHD zinc-finger domains, chromo domains, and an ATPase/helicase domain which is similar to ones found in the SWI2/SNF2 superfamily of proteins⁹. Although no function has yet been ascribed to CHD3 or CHD4, other members of the SWI2/SNF2-related proteins are known to be involved in several ATP-dependent chromatin regulatory processes, including transcriptional activation and

repression, DNA repair and chromosome recombination². HDAC and CHD immune complexes were incubated with the radiolabelled and photoreactive ATP analogue 8-azido- γ -³²P-ATP to test for ATP binding. Irradiation of the complexes with ultraviolet light resulted in the specific crosslinking of ATP with a polypeptide of relative molecular mass 240K, which corresponds to that of the CHD3 and CHD4 proteins (Fig. 4a). The same ATP-binding activity is present in both anti-HDAC as well as in anti-CHD complexes and is competitively inhibited by a 1,000-fold excess of unlabelled ATP.

We next investigated whether CHD3/4- and HDAC-containing NRD complexes could use the energy of ATP hydrolysis to remodel chromatin, as we have observed nucleosome-stimulated ATPase activity from the NRD complex (J.K.T., C.A.H. and S.L.S., unpublished results). Reconstituted, rotationally phased mononucleosome cores were used as a substrate to assay for any changes that the NRD complex might impart on the structure of the nucleosome as reflected by an altered sensitivity of the DNA to either DNase I or *Pst*I restriction endonuclease. Mononucleosome cores were treated with varying concentrations of NRD complex from either HDAC or CHD immunoprecipitates, with or without ATP, and then digested with DNase I. Whereas DNA from control nucleosome cores yields a repeating ~10-base-pair (bp) pattern after digestion with DNase I, addition of both anti-HDAC and anti-CHD NRD complexes disrupts this pattern in a dose- and ATP-dependent manner (Fig. 4b). The pattern of DNase I hypersensitivity that results from treatment

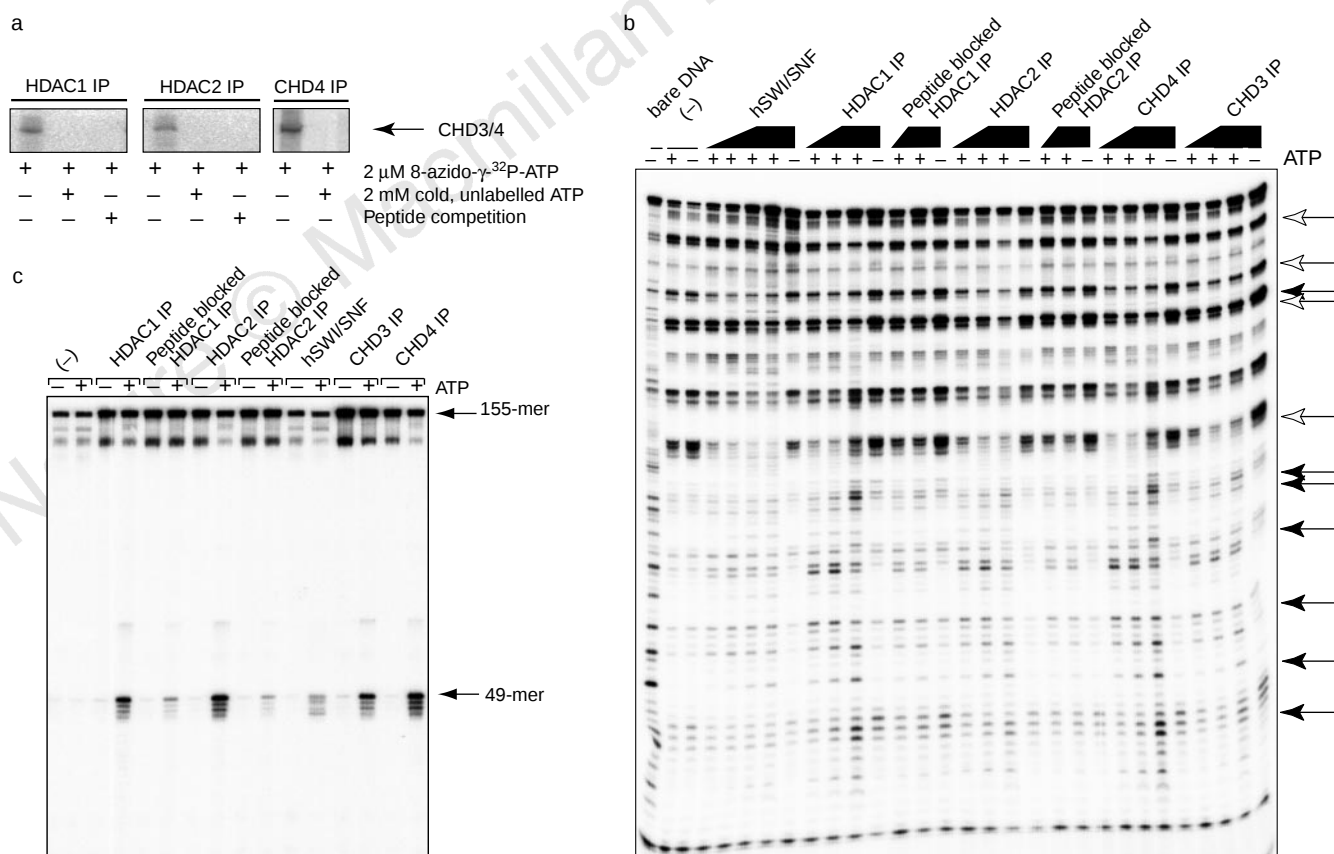


Figure 4 Functional characterization of CHD3/4. **a**, NRD complexes contain a 240K ATP-binding protein(s). NRD complexes prepared from HDAC or CHD immunoprecipitates were crosslinked with a radiolabelled, photoreactive ATP analogue, 8-azido-ATP by irradiation with ultraviolet light. The radiolabelled 240K band is specific to the NRD complex: excess, unlabelled ATP specifically competes with radiolabelled ATP for binding. IP, immunoprecipitate. **b**, HDAC- and CHD-containing complexes disrupt mononucleosomes in an ATP- and dose-dependent manner. Black arrowheads indicate bands of increased DNase I cleavage, white arrowheads indicate bands of decreased DNase I cleavage. The

weak disruption by CHD3 is consistent with the small amount of CHD3 immunoprecipitated. **c**, HDAC- and CHD-containing complexes facilitate *Pst*I cutting of DNA in an ATP-dependent manner. HDAC and CHD immunoprecipitates were incubated with 2.5 ng of mononucleosome cores in the presence or absence of 1 mM ATP/MgCl₂. Reactions were treated with 20 U of *Pst*I for 1 h and separated by gel electrophoresis. Percentage cutting (cut/[cut + uncut]) was determined by phosphorimager analysis. The triplets result from incomplete Klenow fill-in during labelling.

of mononucleosome cores with the NRD complex is similar to that obtained after treatment of mononucleosome cores with the human SWI/SNF complex. There is a background ATP-stimulated remodelling activity that is at least 25-fold less than that found in the NRD and human SWI/SNF complexes. We assayed mononucleosome cores under similar reaction conditions for a 49-bp product after cleavage by *Pst*I. Formation of this product was stimulated between 10- and 20-fold by ATP in the presence of NRD complex from either HDAC1/2 or CHD3/4 immuno-

precipitates, as measured by phosphorimager quantification (Fig. 4c). Again, background ATP-stimulated activity was significantly less than that in the NRD and hSWI/SNF complexes. Because of the similarity in remodelling activities between the NRD complex and hSWI/SNF, we tested the NRD complex for the presence of the BRG1 catalytic subunit of hSWI/SNF. The trace amount (<0.4 ng) of BRG1 that associates with protein A-agarose beads, if present in intact hSWI/SNF complexes, is insufficient to generate the disruption we observed (data not shown). We conclude that the NRD complex possesses a new nucleosome-remodelling factor that is distinct from that of hSWI/SNF. Our ATP crosslinking data indicate that this remodelling activity is derived from the CHD3/4 proteins.

The association of both a histone deacetylase activity and a nucleosome remodelling activity in the NRD complex suggests a co-dependence model in which one activity prepares the substrate for the other. Oligonucleosomes in the absence of ATP are ineffective substrates for HDAC immune complexes *in vitro*, whereas mononucleosomes are better substrates⁴. Using a reconstituted linear-plasmid chromatin substrate, we tested whether ATP-dependent nucleosome remodelling could help the deacetylation of histones by HDAC. In deacetylation reactions containing the NRD complex and plasmid chromatin, we observed enhanced histone deacetylation that was dependent on ATP (Fig. 5a). This enhanced deacetylation was sensitive to treatment with trapoxin, indicating that deacetylation was by HDAC and not by any other factor. The background deacetylation in the absence of ATP could be explained by deacetylation of more accessible nucleosomes at the DNA ends. There was no ATP-stimulated deacetylation in similar reactions using histones, or histones plus DNA (non-nucleosomal) as substrates, consistent with the idea that ATP hydrolysis supplies the energy to overcome structural barriers to deacetylation that are inherent in nucleosomes but not in free histones (Fig. 5b). Incubation of the NRD complex with trapoxin does not prevent this complex from disrupting the DNase I digestion pattern of mononucleosomes (Fig. 5c), indicating that chromatin remodelling occurs independently of histone deacetylation.

Genetic and biochemical evidence supports a role for both histone acetylation/deacetylation and nucleosome remodelling in altering chromatin structure and gene transcription^{10,11}. We have shown that a histone deacetylase exists with a nucleosome-remodelling factor in a functionally coupled complex. The HDAC-CHD interaction is conserved among metazoans in what is probably an evolutionarily important association⁵. Heterochromatin co-localizes with hypoacetylated histones, and certain chromatin domain-containing proteins, including HP1 and Polycomb, are components of heterochromatin¹. The presence of chromatin domains in CHD proteins raises the possibility that HDACs are targeted to heterochromatin through their interaction with CHD3/4. That HDACs function at heterochromatin has previously been suggested by genetic experiments in *Drosophila* and yeast¹². It remains to be investigated what role associated histone deacetylases and nucleosome remodelling activities have in transcriptional regulation and chromatin structure.

Methods

Large-scale purification of the HDAC2 complex. Cell extract from 2.4 litres of HeLa cells (0.5×10^6 per ml) was prepared⁴. Proteins were immunoprecipitated using affinity-purified anti-HDAC2 antibody-protein A beads as described⁴ and eluted by addition of HDAC2 immunogenic peptide. Proteins were re-precipitated in 6% trichloroacetic acid/0.04% deoxycholate and then separated by SDS-PAGE and stained with Colloidal blue (Novex). Bands corresponding to p70, p110 and p240 were excised and samples were sequenced by peptide mass spectrometry.

Immunoprecipitation and western blotting. HeLa and Jurkat cell extracts were prepared as described⁴. 1.25, 5 and 7.5 mg cell extract were immunoprecipitated for western blotting, azido-ATP crosslinking, and mononucleosome disruption assays, respectively. For mononucleosome disruption, immunopre-

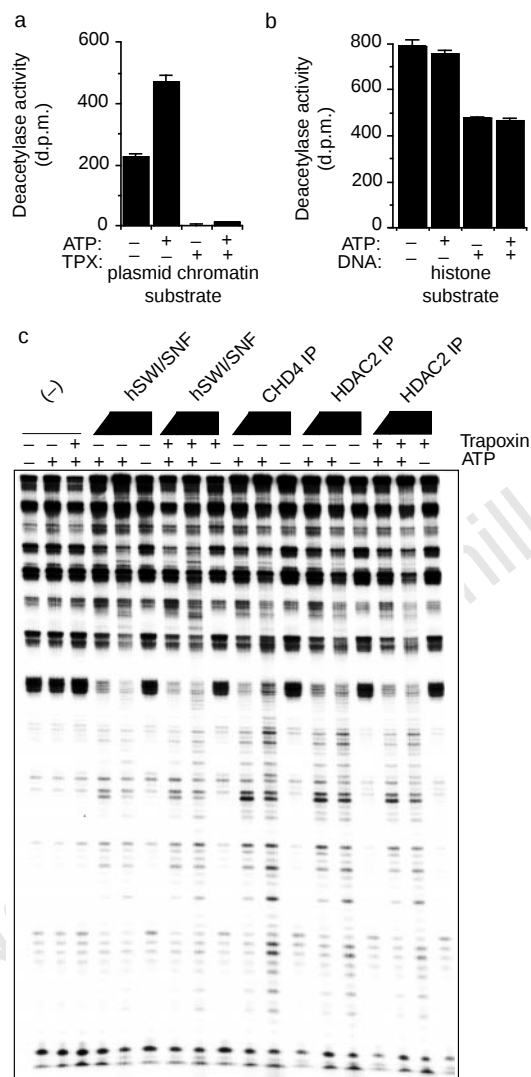


Figure 5 Functional interdependence of histone acetylation and nucleosome remodelling. **a**, Deacetylation of plasmid chromatin histones is aided by ATP-dependent nucleosome remodelling. Chromatinized histones were treated with NRD complex prepared from HDAC or CHD immunoprecipitates in the presence or absence of ATP and the extent of deacetylation was determined by scintillation counting. Treatment of the same reactions with trapoxin reduces deacetylation to background, irrespective of the presence of ATP. A background of 37 d.p.m. has been subtracted. **b**, Deacetylation of non-chromatinized substrates is not dependent on nucleosome remodelling. The same histones and DNA that were used to reconstitute the plasmid chromatin as in (**a**) were incubated with the NRD complex as either histones alone or histones plus DNA under conditions in which nucleosomes do not form. No ATP-stimulated deacetylation occurred. DNA is weakly inhibitory, possibly because of nonspecific interaction with histones. **c**, Remodelling of mononucleosome cores does not depend on histone deacetylase. Cores were treated as in (**b**) in the presence or absence of 200 nM trapoxin and digested with DNase I. The remodelling activity of these complexes was not affected by inhibition of histone deacetylase.

cipitated beads were resuspended in 23 μ l lysis buffer and aliquoted as indicated.

Histone deacetylase assays. Assays were done as described¹; where indicated, ATP was added to 2.5 mM in the presence of 5 mM MgCl₂, and trapoxin was added to 200 nM.

8-azido-adenosine- γ -³²P-ATP crosslinking. NRD antibody-immobilized bead complexes were incubated with 2 μ Ci 8-azido-adenosine-5'- γ -³²P-triphosphate (ICN) for 3 min at room temperature in 50 μ l buffer containing 20 mM HEPES, pH 7.6, 100 mM KCl, 5 mM MgCl₂, 1 μ M ZnSO₄, 0.1% Tween-20. The reaction mixture was irradiated for 5 min with 254-nm UV light (1,120 μ W cm⁻² from 3 cm away). Beads were washed three times with 1 ml 20 mM HEPES, pH 7.6, 100 mM KCl, 0.1% Tween-20. Bound proteins were resolved by SDS-PAGE and imaged using a phosphorimager.

Nucleosome-disruption assays. the 155-bp pTPT *MluI/EcoRI* fragment was prepared and footprinted as described¹³. 0.2, 1.0 and 5.0 μ l HDAC/CHD immunoprecipitates, 1.0 and 5.0 μ l peptide-blocked HDAC/CHD immunoprecipitates, and 0.04, 0.16, 0.6 and 2.4 μ l of hSWI/SNF were tested for disruptive activity. 2 mM ATP/MgCl₂ was added where indicated; after 45 min, reactions were treated with 0.12 U DNase I (0.012 U for bare DNA).

Plasmid chromatin reconstitution. HeLa cells were radiolabelled as described previously⁴ and hyperacetylated histones were purified by hydroxyapatite chromatography as described¹⁴. 16 μ g of a 9-kb plasmid was linearized, purified, and mixed with an equal mass of radiolabelled, hyperacetylated core histones in 50 μ l of 15 mM HEPES, pH 7.5, 1 M NaCl, 0.2 mM EDTA and 0.2 mM PMSE. Stepwise dilution of salt was done by addition of the same buffer without NaCl over 3 h at 10-min intervals until the final concentration of NaCl was 100 mM. Non-nucleosomal histones were then separated by repeat Centricon-500 (Amicon) filtration. Nucleosomal structure was confirmed by micrococcal nuclease digestion of the product.

Received 24 August; accepted 28 September 1998.

1. Elgin, S. C. R. *Chromatin Structure and Gene Expression* (eds Hames, B. D. & Glover, D. M.) (IRL, Oxford, 1995).
2. Pazin, M. J. & Kadonaga, J. T. SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein-DNA interactions? *Cell* **88**, 737–740 (1997).
3. Imhof, A. & Wolffe, A. P. Transcription: gene control by targeted histone acetylation. *Curr. Biol.* **8**, R422–424 (1998).
4. Hassig, C. A. *et al.* A role for histone deacetylase activity in HDAC1-mediated transcriptional repression. *Proc. Natl Acad. Sci. USA* **95**, 3519–3524 (1998).
5. Wade, P. A., Jones, P. L., Vermaak, D. & Wolffe, A. P. A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. *Curr. Biol.* **8**, 843–846 (1998).
6. Tavladoraki, P. *et al.* Maize polyamine oxidase—primary structure from protein and cDNA sequencing. *FEBS Lett.* **426**, 62–66 (1998).
7. Seelig, H. P. *et al.* The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation. *Arthr. Rheum.* **38**, 1389–1399 (1995).
8. Ge, Q., Nilasena, D. S., O'Brien, C. A., Frank, M. B. & Targoff, I. N. Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2 autoantigen. *J. Clin. Invest.* **96**, 1730–1737 (1995).
9. Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P. & Collins, F. S. Characterization of the CHD family of proteins. *Proc. Natl Acad. Sci. USA* **94**, 11472–11477 (1997).
10. Hartzog, G. A. & Winston, F. Nucleosomes and transcription: recent lessons from genetics. *Curr. Opin. Genet. Dev.* **7**, 192–198 (1997).
11. Kadonaga, J. T. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* **92**, 307–313 (1998).

12. De Rubertis, F. *et al.* The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* **384**, 589–591 (1996).
13. Schnitzler, G., Sif, S. & Kingston, R. E. Human SWI/SNF interconverts a nucleosome between its base state and a stable remodeled state. *Cell* **94**, 17–27 (1998).
14. Cote, J., Utley, R. T. & Workman, J. L. Basic analysis of transcription factor binding to nucleosomes. *Meth. Mol. Genet.* **6**, 108–128 (1995).

Acknowledgements. We thank W. Wang for unpublished CHD3 and CHD4 antisera and for sharing unpublished results; I. Targoff for human Mi-2-positive antiserum and rabbit anti-Mi-2 α antiserum; W. Lane and colleagues at the Harvard Microsequencing Facility for peptide microsequencing; O. Rando, C. Grozinger, D. Ayer, and I. Wilson for discussion; and the NIH Cell Culture Center for technical assistance. This work was supported by a National Institute of General Medical Sciences grant to S.L.S. and an NIH grant to R.E.K. G.R.S. is a Helen Hay Whitney Fellow. Predoctoral fellowships from the NSF and the Harvard-Markey Biomedical Scientist Program to J.K.T. and an NIH predoctoral training grant to C.A.H. are gratefully acknowledged. S.L.S. is an Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to S.L.S. (e-mail: sls@slsiris.harvard.edu).

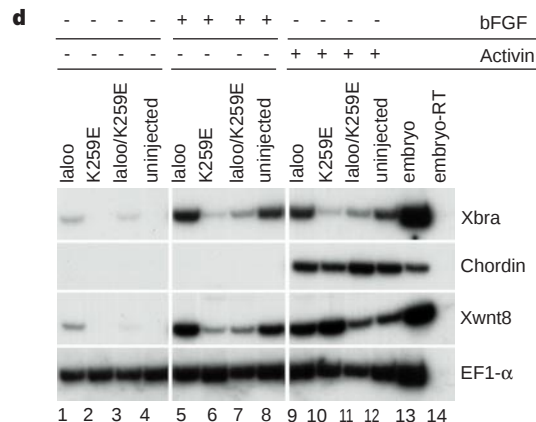
erratum

FGF-mediated mesoderm induction involves the Src-family kinase Laloo

Daniel C. Weinstein, Jennifer Marden, Francesca Carnevali & Ali Hemmati-Brivanlou

Nature **394**, 904–908 (1998)

In Fig. 4d of this Letter, it was wrongly indicated that activin was added to samples 5–8. Activin was not added to these samples but to samples 9–12 as shown here.



Neuroscience on the brain

The pilgrimage to the annual Society for Neuroscience meeting will convene in Los Angeles from the 7th through the 12th of November – readers may save their legs by targetting some of the items featured here.

Current Protocols in Neuroscience

From John Wiley & Sons

An up-to-date collection of neuroscience methods

The publication draws from techniques in neurophysiology, neuroanatomy, neuropharmacology and behavioural neuroscience to meet the needs of researchers in the full range of disciplines involved in studying the brain, nervous system, and corresponding behaviours. According to Wiley, *Current Protocols in Neuroscience* features carefully edited techniques with troubleshooting tips and helpful comments that come from extensive experience in using these protocols. Each feature provide step-by-step guide through methods. Moreover, quarterly updates are available in looseleaf and electronic formats and should help to keep laboratories up to date with the latest developments in this rapidly changing field. Forthcoming coverage includes gene expression in transgenic mice using neural promoters, generation and utilization of phosphorylation state-specific antibodies to investigate neuronal signaling pathways, an overview of the baculovirus expression system and the measurement of chloride movement in neuronal preparations, to name a few.

Reader Enquiry No. 100

Human prion Pr²⁷⁻³⁰ peptide pAb

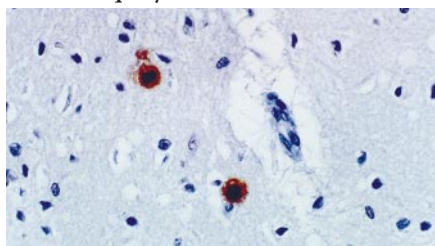
From Chemicon

www.chemicon.com

A goat polyclonal antibody to the human, prion PrP₂₇₋₃₀ peptide

This antibody is made against a synthetic peptide that corresponds to amino acids 79–97 of the amino terminus of the human prion protein PrP₂₇₋₃₀. This has been shown to be immunoreactive to the peptide by ELISA. Moreover, the antibody will immunolabel amyloid plaques in formalin-fixed paraffin sections from CJD brain sections. This antibody should prove useful to scientists and neuropathologists who study the pathogenesis of prion diseases.

Reader Enquiry No. 101



Getting your goat, polyclonal antibody, that is, to human prion peptide PrP₂₇₋₃₀.

NT-4 E_{max} immunoassay

From Promega

www.euro.promega.com/uk

A system for sensitive and quantitative detection of neurotrophic factor-4 (NT-4) in an antibody sandwich format

Soluble NT-4 is captured with an anti-human, NT-4 polyclonal antibody and then binds a second, specific anti-NT-4 monoclonal antibody. This monoclonal antibody is detected using a species-specific antibody conjugated to horseradish peroxidase. The antibodies used in this assay provide specific detection of NT-4 to a minimum of 9.4 pg ml⁻¹ and demonstrate less than three per cent crossreactivity with other related neurotrophic factors. The immunoassay system can be used to quantitate NT-4 in tissue culture supernatants, plasma, serum and tissue extracts. Available in two sizes, with two or five 96-well ELISA plates, this system can be configured as desired. Other immunoassays in this range include TGFβ1 and TGFβ2, BDNF, GDNF, gp130, NGF and NT-3.

Reader Enquiry No. 102

Reagents for Alzheimer's research

From Serotec

www.serotec.co.uk

Human polyclonal antibodies and recombinant proteins for researching Alzheimer's disease

Products currently available from the company for this area of research include presenilin-1, amyloid-β, amyloid-β precursor and apolipoprotein E. Many of these products are suitable for use in a wide range of applications, including immunocytochemistry, ELISA, immunoprecipitation and western blotting.

Reader Enquiry No. 103

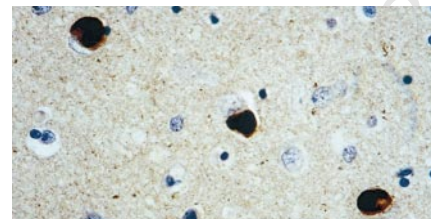
Synuclein antibodies

From Affiniti Research Products

www.affiniti-res.com

Inclusions known as Lewy bodies may contribute significantly to neuronal malfunction and finally to cell death

Found in Parkinson's disease, Lewy bodies are known to contain accumulations of α-synuclein, ubiquitin, and various proteinaceous elements of the ubiquitin-proteasome pathway. Polyclonal antibodies specific to α- and β-synuclein have recently been developed and added to the company's listing of products for brain disease research. The peptide-derived antibodies are suitable for use in western blotting and immunohistochemical applications. In



New synuclein antibodies offered by Affiniti.

addition to the synuclein antibodies, Affiniti offers some 40 antibodies to proteasome subunits and ubiquitin-related compounds, together with a comprehensive range of substrates and proteasome inhibitors. Also offered are antibodies to neurofilaments, tau, tau-protein kinase (GSK3β), and other antigens that are of importance to understanding the mechanisms of neurodegeneration.

Reader Enquiry No. 104

Mdr1a/b double knockout mouse

From Taconic

www.taconic.com

A new double knockout mouse is offered for research into the central nervous system

This mouse contains a disruption of multi-drug resistance genes mdr1a and mdr1b, it has a functional deficiency in the blood brain barrier and is applicable to a wide range of central nervous system research. It should prove useful for studies involving neurotoxicology, drug transport, chemotherapy, oral bioavailability and multi-drug resistance. The mdr1a/b double knockout mouse was generated by sequential gene targeting of embryonic stem (ES) cells in the laboratory of Alfred Schinkel *et al.* at the Netherlands Cancer Institute. According to Taconic, these mice exhibit normal development, viability and fertility. Moreover, the mdr1a/b double knockout model has shown an increase in the accumulation of digoxin in the brain and testis (male) — this is similar to the mdr1a knockout mouse model. The mdr1a/b mouse also showed a significant increase in accumulation of digoxin in the ovaries and adrenal glands, over plasma levels and over



Double trouble? Not with this double knockout.

levels in wild type mice, states Taconic. The *mdr1a/b* knockout mouse is available on a FVB background.

Reader Enquiry No. 105

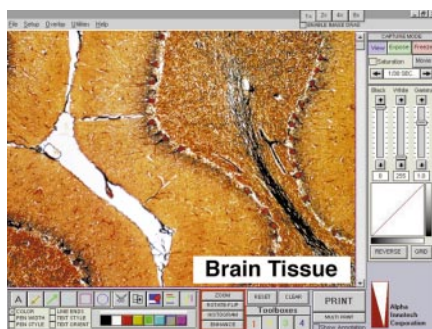
Imaging systems

AlphaEase 4.0 software

From Alpha Innotech www.alphainnotech.com

Imaging software for the AlphaImager and ChemImager systems

Centred on 32-bit architecture, this program supports long file names, enhanced dialogue boxes, and a more familiar look and feel to the software features. Features include MovieMode, camera exposure control from 0.033 s to 4 h, nine levels of automatic image enhancement, zoom/pan controls, histograms, filters and false colours. Analysis tools include one-dimensional densitometry, molecular weights, high-density array analysis, one-dimensional multi-lane analysis, two-dimensional spot densitometry analysis, manual and automatic count for colonies and cells, gel scoring and ruler functions. This new architecture yields more robust compatibility with most Windows applications, as well as enhanced multitasking. AlphaEase 4.0 also boasts new updates that



AlphaEase imaging software from AlphaInnotech.

include 8- and 16-bit grayscale file import, 24-bit colour, higher resolution file import and TWAIN acquisition.

Reader Enquiry No. 106

New technical notes

From Bio-Rad

www.bio-rad.com

Two new notes that cover multi-photon imaging of biological tissue

Technical note five summarizes the benefits of non-descanned (external) detectors in collecting scattered light and improving sensitivity, allowing deeper optical sectioning and detection of faint fluorescent signals.

It also illustrates the design of the external detector on the MRC-1024MP system and lists the full range of available filters. On the same subject, technical note number seven describes the advantages of using femtosecond, pulsed lasers for achieving deeper imaging into tissue with flexibility in the use of available power. Multi-photon imaging is increasingly used for studying biological tissue as it allows deeper imaging into samples than other methods and enables the study of living samples for longer periods of time than previously possible, says Bio-Rad.

Reader Enquiry No. 107

PhoTurbo

From Life Science Resources

www.lsr.co.uk

Fluorescence detection for simultaneous real-time photometry and electrophysiology

According to the manufacturer, PhoTurbo provides the very high temporal resolution in data acquisition for experiments that the latest optical probes in high speed neurobiological research applications. It has two-in-one functionality, as the system is configured with both PhoCalTurbo and PhoClampTurbo on a single platform. These systems enable both chart-recorder-like data acquisition and triggered sweep acquisition, with full experimental control and on-line viewing of data. PhoTurbo has 16-bit precision and dynamic range to acquire both fluorescence and analog data at rates up to 64 kHz. According to the manufacturer, high-sensitivity photon-detection measurement makes it possible to record and display online even the fastest of physiological events in real time. The system measures optical probes used in the detection of membrane potential or ions, such as calcium sodium and magnesium, which can be combined with such analog patch clamp measurements as voltage and current, or with extracellular pH, O₂ tension and temperature. PhoTurbo is a complete, integrated system, operating under Windows95 with full 32-bit functionality. It includes photomultiplier tubes, all associated electronics, acquisition hardware, power supplies and workstation.

Reader Enquiry No. 108

DP10 digital camera

From Olympus

www.olympus-europa.com

A digital photomicrography system for high performance at a more reasonable price

The DP10 digital, colour camera has a progressive 1.7-cm CCD sensor with over 1.4 million pixels (1,280 × 1,024 pixels). However, image resolution is not artificially increased or interpolated by software as is common with other cameras within this price range. Instead, every pixel of the CCD sensor is represented directly in the digital picture. The technology is said to ensure that pictures are sharp with natural colour

ADVERTISEMENT

A Super Spectrum of Possibilities

Pierce SuperSignal® Chemiluminescent Substrates for all your protein research needs.

Western Blotting Applications

SuperSignal® West Pico Chemiluminescent Substrate*
Picogram-level sensitivity at an amazingly economical price
Product # 34080EL, 500 ml
(Formerly SuperSignal® Chemiluminescent Substrate for Western Blotting)

SuperSignal® West Dura Extended Duration Substrate*
24-hour signal duration and super sensitivity
Product # 34075EL, 100 ml
(Formerly SuperSignal® ULTRA Chemiluminescent Substrate)

SuperSignal® West Femto Maximum Sensitivity Substrate*
The most intense signal and the greatest sensitivity attainable with a chemiluminescent substrate
Product # 34085EL, 100 ml
(Formerly SuperSignal® BLAZE™ Chemiluminescent Substrate)

ELISA

SuperSignal® ELISA Pico Chemiluminescent Substrate*
Picogram-level detection for plate-based assays
Product # 37070EL, 100 ml
(Formerly SuperSignal® LBA for Luminometer-Based Applications)

SuperSignal® ELISA Femto Maximum Sensitivity Substrate*
A more sensitive substrate with enhanced low-end linearity
Product # 37075EL, 100 ml
(Formerly PowerSignal® ELISA Chemiluminescent Substrate)
*Pierce Patent

Call for your FREE Substrates Selection Guide Poster!
Call 800-874-3723 for product information or 815-968-0747 for the name of your local distributor.

PIERCE
A Thermo Life Sciences Company

3747 N. Meridian Rd. • PO Box 117 • Rockford, IL 61135 U.S.A. • Fax 815-968-7216
E-mail: Thermo@piercelab.com • Internet: <http://www.piercelab.com> © Pierce Chemical Company, 1998

READER ENQUIRY NO. 109